Methodology of Educational Research

Lokesh Koul

Knowledge of research methodology is essential to conduct research projects and to keep oneself abreast of the new developments in education.

An uptodate account of the methods and techniques suited to the field of education, the book provides an understanding of the significant educational problems that need to be tackled. It also discusses the steps that a researcher is expected to take in conducting a research study, the descriptive and inferential statistical techniques; the reporting of educational research; characteristics of historical method, descriptive method and experimental method. Finally, uptodate list of research reports and statistical tables are provided.

Designed and written mainly for students of M.A. (Education, Psychology and Sociology), M Ed. and M.Phil. (Education, Psychology and Sociology), the book will be of immense value to Ph.D. students and other researchers of Social and Biological Sciences.

1V2KA821

Rs 295

Methodology of Educational Research

LOKESH KOUL
Professor, and Dean
Faculty of Education
Himachal Pradesh University
SHIMLA

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

VIKAS PUBLISHING HOUSE PVT LTD

VIKAS PUBLISHING HOUSE PVT LTD
576 Masjid Road, Jangpura, New Delhi-110014.

COPYRIGHT © LOKESH KOUL, 1984

First Edition, 1984
Second Revised Edition 1988

370.78
KOU

C.E.R West Bengal
ate 29.1.91
cc. No 4849

All rights reserved. No part of this publication may be reproduced in any form without the prior written permission of the Publishers.

To
Saint Dina Nathji Razdan and
my parents Shrimati Gunawati and
Shri S.K. Koul with profound gratitude
for their blessings

Printed at : K.P.S., Shahdara, Delhi-110032

Preface to the Second Revised Edition

The first edition of the book was published in 1984. The book has been very popular among the students, research scholars and teachers of Education, Psychology, Sociology and other Social Sciences at the Post-Graduate, M. Phil. and Ph. D. degree levels. The presentation of the content material with examples and illustrations drawn from Indian conditions in each Chapter beginning with its brief exposition along with a set of learning objectives, and followed by a post-test at the end has been widely acclaimed by the readers as distinctively unique feature of the book.

The second edition has been revised and updated keeping in view the priority areas of research in the perspective of New Education Policy—1986. The area of Non-formal Education highlighting the issues related to its implementation as an alternative strategy in the universalisation and furtherance of education has been added in the Chapter—Areas of Educational Research. The Regression Analysis, an important multivariate statistical technique along with suitable illustrations and the Content Analysis with significant modifications, have also been incorporated.

The book in its totality including the additions, alterations and modifications carried out in the present revised edition is expected to be of added help : (i) in the understanding and appreciation of the role of research in the theory and practice of education; (ii) to develop competence to plan, execute and report research; and (iii) in enhancing the ability to apply research findings and to comprehend the essential of reported research.

LOKESH KOUL

Preface to the First Edition

Research in Education, as in other fields, is a search for knowledge. It is not a search that yields infallible truths; but it is rather a search that provides knowledge for the solution of problems in the field of education. A knowledge of the research methodology, therefore, is essential for all those who either take an active role in the conduct of research or desire to keep themselves abreast of the new developments in education.

This book has been written as a basic book covering various aspects of educational research. It may also serve as a useful reference for classroom teachers, students working for their doctoral degree, or administrators who are interested in conducting research in education. The book is designed to provide not only the knowledge of research methodology but also an understanding of the significant educational problems that need to be tackled in our country. All important aspects of educational research have been included both for *producer* of research concerned with the growth of education as science, and for the *consumer* of research interested in the interpretation and application of research findings.

A distinctive feature of the book is that each chapter begins with a brief exposition of its content. It is then followed by a set of learning objectives which the reader is expected to meet upon completing the chapter. At the end of each chapter a post-test is provided. It helps to determine how well a reader has grasped the material and what objectives require that he go through that relevant parts of the chapter once again.

Chapter 1 outlines the conceptual framework which is essential for understanding the meaning and characteristics of research. It next explains basic, applied and action research. The nature, historical development, and limitations of educational research are other principal elements dealt with in this chapter.

Chapter 2 presents a discussion of the areas of educational research relating to the content of education; stages and sectors of education; and teaching process. Some priority areas of educational research in India have also been pinpointed in this chapter.

The sequence of the next four chapters roughly follows the steps that a researcher might be expected to take in conducting a research study. These include the techniques of problem identification, hypothesis formulation, location and review of related literature, methods of sample selection, the design and use of data gathering tools, data analysis techniques including the use of computational aids, and interpreting research findings. Although it is not the purpose of this book to provide expertise in the mechanics of modern computational aids like desk calculators and electronic digital computers, it attempts to make the reader aware of their existence and utility.

Chapters 7 and 8 discuss in detail the descriptive and inferential statistical techniques which are commonly used in the analysis and interpretation of data. The reporting of research results and findings, which marks the final stage in the research process, is discussed in *Chapter 9*. The last three chapters embody a detailed discussion of three common research methods: historical, descriptive, and experimental. The salient feature of the last chapter, on experimental method, is the inclusion of a detailed discussion on experimental designs with illustrations.

I am grateful to many of my colleagues and research students in the Faculty of Education, Himachal Pradesh University, Shimla. I am also indebted to various authors whose books and articles I consulted during the preparation of this book.

I have no words to express my gratitude to my wife Prabha who gave me every sort of cooperation and inspiration and sustained me during the long period of this work's preparation.

LOKESH KOUL

Contents

1. INTRODUCTION 1—23

Methods of Acquiring Knowledge; The Nature of Science; The Meaning and Characteristics of Research; The Nature of Educational Research; Basic and Applied Research

2. AREAS OF EDUCATIONAL RESEARCH 24—56

Problems Relating to Content of Education; Problems Relating to Different Stages and Sectors of Education; Problems Relating to Teaching Process; Priority Areas of Educational Research in India

3. THE RESEARCH PROBLEM AND PREPARATION OF THE RESEARCH PROPOSAL 57—82

Knowing the Field; Selection of the Problem; Definition and Statement of the Problem; Evaluation of the Problem; The Hypotheses; The Research Proposal or Synopsis; Introduction; Procedures for Collecting Data; Procedures for Treating Data; Bibliography; Time Schedule; Budget Schedule

4. REVIEW OF THE RELATED LITERATURE 83—103

Purpose of the Review; Identification of the Related Literature; Organizing the Related Literature

5. DATA COLLECTION 104—183

Sampling; Methods of Sampling; Data-Gathering Tools; Psychological Tests; Inquiry Forms; Observation; Interview; Sociometric Techniques

6. ANALYSIS AND INTERPRETATION OF DATA 184—196

Organization of Data; Analysis of Data; Modern Computational Mechanical Aids; Language of the Computer; Elements of a Computer System; Interpretation of Data; Formulation of Conclusions and Generalizations

7. STATISTICAL METHODS 197—272

Measurement and its Levels; Types of Data; Classification of Data; Graphic Representation of Data; Descriptive Statistical Measures; Measures of Central Tendency; Measures of Variability; Normal Probability Curve; Measures of Relative Positions; Measures of Relationship; Uses of Correlation; Limitation of Statistical Methods

8. INFERENTIAL STATISTICS 273—351

Statistical Inference Based on Parametric Tests; Statistical Inference Based on Non-Parametric Tests

9. THE RESEARCH REPORT 352—383

General Format of the Research Report; Style and Format of Writing; Typing of the Research Report; Proof Reading of the Final Draft of the Research Report

10. METHODS OF EDUCATIONAL RESEARCH 384—399

The Historical Method; Nature of Historical Research; Value of Historical Research in Education; Types of Historical Research; Steps in Historical Research

11. ✓THE DESCRIPTIVE METHOD 400—438

Nature of Descriptive Research; Value of Descriptive Research in Education; Types of Descriptive Research; Steps in Descriptive Research

12. ✓ EXPERIMENTAL METHOD 439—483

Nature of Experimental Research; Steps in Experimental Research; Experimental Designs; Types of Experimental Designs

REFERENCES 484—493

APPENDICES 494—506

1. Introduction

The general purposes of this chapter are reflected in four main sections. The first section describes the various methods of acquiring knowledge, the nature of scientific inquiry and the steps in scientific method. The second section explains the nature of science, the role of scientific theories in the development of science and the characteristics of a sound theory. The third section deals with the meaning and characteristics of research. The fourth section describes the nature, historical development and limitations of educational research, the need for research in education, and the steps in educational research. The last section compares the basic research with the applied and action research.

Learning Objectives

After carefully reading this chapter you will be able to meet the following objectives :

1. Describe the various methods of acquiring knowledge.
2. Describe the nature of scientific method.
3. List and describe the steps used in scientific method.
4. Describe the nature of science.
5. Describe the role of scientific theories in the development of science.
6. List and describe the characteristics of a sound theory.
7. Describe the meaning of research.
8. Describe the characteristics of research.
9. Describe the nature of educational research.
10. Describe the need for research in education.
11. Describe the stages in the development of educational research.
12. List and describe the limitations of educational research.
13. List and describe the steps in educational research.
14. Define basic, applied and action research.
15. Distinguish between basic, applied and action research.

Human knowledge as it exists today broadly consists of facts and theories. New facts, new concepts and new ways of doing things increased its quantum with the passage of time. This knowledge enables us to understand, comprehend, explain, control, predict, or cope with a given situation. The sources from which we obtain knowledge range from those that are highly reliable to those that are completely unreliable. The knowledge obtained from the unreliable sources is based on assumptions, beliefs, and untested generalizations.

which one proceeds from general to specific statements by *deductive reasoning*. It provides a means of testing the validity of any given conclusion or idea by proceeding from the known to the unknown. The syllogistic reasoning consists of (1) a major premise based on a self-evident truth or previously established fact or relationship; (2) a minor premise concerning a particular case to which the truth, fact, or relationship invariably applies; and (3) a conclusion. If the major and minor premises can be shown to be true, the conclusion arrived at is necessarily true. To use a simple example, consider the following proposition :

1. All animals are mortal;
2. Dog is an animal;
3. Therefore, dog will die.

The method of syllogism or deduction, however useful, has the following limitations:

1. The conclusion of a syllogism is always derived from the content of premises. Therefore, if the permises are unrelated or if one of the premises is erroneous, the conclusion arrived at will not be valid.

2. Another serious limitation of the deductive reasoning is its dependence upon verbal symbolism.

3. Deductive reasoning can systematize what is already known and can identify new relationships as one proceeds from known to unknown, but it cannot be relied upon as a self-sufficient method for securing reliable knowledge.

Inductive Reasoning

The conclusions derived from generalities and from statements of presumed authorities by deductive reasoning are true only if they are based upon true premises. To determine whether the premises are true, Francis Bacon stressed the need for basing general conclusions upon specific facts gathered through direct observations. This is what is known as *inductive reasoning*, that is, going from the particular to the general. Rather than accepting premises laid down by authorities as absolute truths, Bacon advised man to observe nature closely, to experiment, to tabulate all the facts, to study these facts in order to reach minor generalizations, and then to proceed from minor generalizations to greater ones. He, however, cautioned against formulating any hypothesis or any probable solution to a problem until all the facts had been gathered.

In deductive reasoning the premises or generalizations must be known before a conclusion can be reached. On the other hand, in inductive reasoning a conclusion is reached by observing instances and generalizing from instances to the whole phenomenon. In order to be absolutely certain of an inductive conclusion, all instances must be observed. Under Baconian system of reasoning it is known as perfect induction. In practical situations, however, it is not possible to examine every instance of a phenomenon to which a generalization refers. When examining all the instances of a phenomenon under study is not practical, one may arrive at a generalization by observing only some instances that make up the phenomenon. This is known as *imperfect induction.*

Although imperfect induction does not help man to arrive at infallible conclusions, it can provide him reliable knowledge upon which he makes reasonable decisions.

Both inductive and deductive methods have advantages and limitations. If premises are true, deductive reasoning helps to arrive at absolutely true conclusions. These conclusions, however, do not probe beyond that, which is already known—already present, at least implicitly, in the premises. The conclusions reached by imperfect inductive reasoning do contain information that is not present, even implicitly, in one of the premises (the observed instances). If all the premises (observed instances) are true, the probability of conclusions arrived at may be of varying degrees.

The Scientific Method

The exclusive use of Bacon's inductive method resulted in the accumulation of isolated bits of information and therefore it made little contribution to the advancement of human knowledge. Moreover, many problems could not be solved by inductive reasoning alone. In view of these limitations it was superseded by the *inductive-deductive method* or the *scientific method.* This method, generally attributed to Charles Darwin, integrates the most important aspects of the inductive and deductive methods. The scientific method, therefore, is a back-and-forth movement of thought in which man first operates inductively from partially known or sometimes confused information learned from experience, previous knowledge, reflective thinking, observation and so on, towards a meaningful whole or hypothesis, and then deductively from suggested whole or hypothesis to the particular parts in order to connect these with one another in a meaningful pattern to find valid relationships. In the words of Dewey 1933, p. 87):

> While induction moves from fragmentary details (or particulars) to a connected view of situation (universal), deduction begins with the latter and

works back again to particulars, connecting them and binding them together.

Although, in practice, scientific method involes a double movement of reasoning from induction to deduction, in its simplest form, it consists of working inductively from observations to hypotheses and then deductively from the hypotheses to the logical implications of the hypotheses in relation to what is already known.

Steps in the Scientific Method

Scientific method, as pointed out earlier, differs from other methods of acquiring knowledge on the bases of generalizations from authority, tradition, experience, and syllogism. It also differs from the methods of chance, of trial-and-error, and of intuition. When using the scientific method, man engages himself in a thinking process called reflective thinking. The five stages of reflective thinking furnished by Dewey (1911) may be summarized as under (Van Dalen 1973, p. 13; Whitney 1964, p. 3):

1. *The occurrence of a felt difficulty.* Man comes across some obstacle, experience, or problem that puzzles him.
   - (*a*) He lacks the means to achieve the end desired.
   - (*b*) He feels difficulty in identifying the character of an object.
   - (*c*) He is unable to explain an unexpected event.
2. *Identification and definition of the difficulty in terms of a problem statement.* Man makes observations and gathers facts so that he is able to define his difficulty more precisely.
3. *Suggested solutions of the problem-hypotheses.* Man makes intelligent guesses about possible solutions of the problem from the preliminary study of the facts. Such guesses that he makes to explain the facts about the cause of difficulty are called hypotheses.
4. *Deriving consequences of the suggested solutions with the help of deductive reasoning.* With the help of deductive arguments, man reasons that if each hypothesis is true, certain consequences should follow.
5. *Experimental verification of the hypotheses.* Man verifies each hypothesis by searching for observable evidence that will confirm whether or not the consequences that follow actually occur. This process enables him to test which hypothesis is in conformity with observable facts.

The stages involved in reflective thinking presented above suggest a pattern that is employed in the scientific method. It will be seen that the pattern describing this method runs parallel to the stages involved in reflective thinking.

The scientific method also consists of five definite steps:

1. *Identification and definition of the problem.* A scientific inquiry starts with the identification of a problem that is in need of solution. The problem identified must be defined in such a manner that observation or experimentation in the natural world can provide a solution.

2. *Formulation of a hypothesis.* Once the problem is defined, the next step is to formulate the hypothesis, which provides an intelligent guess for the solution of the problem. It requires a critical review of the knowledge or information related to the problem.

3. *Implication of hypothesis through deductive reasoning.* The next step after the formulation of a hypothesis is to deduce the implications of the suggested hypothesis, that is, what would be observed if the hypothesis is true.

4. *Collection and analysis of evidence.* The deduced implications of the hypothesis are tested by collecting relevant evidence related to them through observation, testing, and experimentation.

5. *Verification, rejection, or modification of hypothesis.* Once the evidence has been collected and analysed, the results are analysed in order to verify whether the evidence support the hypothesis. It may be noted that the characteristic of scientific method is not to prove the hypothesis in terms of absolute truth but to conclude that the evidence does or does not support the hypothesis.

THE NATURE OF SCIENCE

From the earlier times man has been curious about anything he could not understand. Slowly and gradually he developed the scientific method of thinking and of investigating his problems which, today is producing astonishing results. It is an orderly system of searching for truth which, by basing conclusions upon factual evidence, and by using reasoning as a means of showing relationship between ideas, has given him better and more accurate answers to his many problems not only in physical and biological sciences but also in behavioural and social sciences. By attempting to apply this method of inquiry to behavioural and social sciences, the fields of psychology, economics, political science, sociology, anthropology, and education have become recognized as sciences. The term *science*, therefore, is now thought of as a method or attitude rather than a field of subject matter. It is described as a method of inquiry that permits man to examine the phenomena of interest to him.

Science is based on certain beliefs and assumptions which are briefly described as under:

1. All events in nature are, at least to a degree, lawful or ordered, predictable and regular. This order, predictability and regularity of nature can be discovered through the activities of the scientific method.

2. Truth can ultimately be derived only from observation. Scientist does not depend upon authority as a source of truth, but relies upon empirical observation. Thus the phenomena that can actually be observed to exist are within the domain of scientific method.

3. The scientist maintains a doubtful attitude towards data. He regards findings as tentative unless they are verified. Verification of the findings requires that other scientists must be able to repeat the observations and get the same results.

4. The scientist is objective, impartial and logical in collecting and interpreting data or making observations. His personal bias does not in any way influence the truth and facts even when they are not in conformity with his own opinions.

5. Scientist does not bother about the moral implications of his findings. He always deals with facts and does not consider what finding is good or what is bad for us.

6. The ultimate goal of science is to integrate and systematize findings into a meaningful pattern or *theory*. The theory, however, is regarded as tentative and not the ultimate truth. It is subject to revision or modification as new evidence is found.

The Role of Scientific Theories

Theories are statements that explain a particular segment of phenomena by specifying certain relationships among variables. According to Kerlinger (1978, p. 9):

> A theory is a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena.

This definition emphasizes three things:

First, a theory is a set of statements comprising of interrelated constructs, definitions and propositions.

Secondly, a theory sets out the relations among the set of constructs, definitions or propositions in order to present a systematic view of the phenomena.

Thirdly, a theory explains a phenomena by specifying what variables are related to what variables and how they are related, thus enabling the scientist to predict from certain variables to certain other variables.

A teacher, for example, has a theory of low academic achievement. His variables might be home environment, intelligence, verbal and numerical aptitudes, anxiety, economic factors and achievement motivation. The phenomenon to be explained is low academic achievement. It is explained by specified relations between each of the seven variables and low academic achievement, or by combinations of the seven variables and low academic achievement. The teacher seeks to understand low academic achievement with the help of this set of relations or constructs. He is then able to 'explain' and to some extent at least 'predict' it. Moreover, he can also to some extent 'control' the low academic achievement by making changes in the environment or manipulating some of the variables.

Scientific theories serve as means and ends in the development of science. As means, they provide a framework which guide scientists in making observation and discovery. Theories summarize knowledge and puts in order the knowledge within a given field. They also clarify and provide meaning to this summarized knowledge comprising of isolated empirical findings. As ends, theories provide scientists explanations for observed events and relationships for specific phenomena with maximum objectivity. They do so by showing what variables are related and how they are related. On the basis of such relationships, scientists make deductions and predict about what will happen in certain situations under specified conditions. In this way theories help in the development of new knowledge.

Characteristics of a Sound Theory

A sound theory must meet the following criteria if it is to serve its purpose in the development of science.

1. A theory must be stated in simple and precise terms. A theory that explains the most in the simple form is preferred to one that has more complexities and assumptions. This is the *Law of Parismony*.

2. A theory must be in conformity both with the observed and with the previously established body of knowledge or validated theories.

3. A theory must provide means for its own interpretation and verification. In other words, it must provide deductions which can be tested empirically.

4. A theory must guide new discoveries and identify areas which are in need of investigation. It can do so if it is based on empirical facts and relationships.

Theories often offer a crude and general explanation of phenomena. They are refined and modified as knowledge in the form of facts accumulates. The discovery of pertinent facts is essential in order to determine whether a theory can be confirmed or should be rejected or reformulated. For example, if the facts found do not substantiate the theory, a scientist must reject or reformulate the theory to fit the new facts.

THE MEANING AND CHARACTERISTICS OF RESEARCH

Research may be defined as the application of the scientific method in the study of problems. At moments the terms *research* and *scientific method* are used interchangeably.

Research is a systematic attempt to obtain answers to meaningful questions about phenomena or events through the application of scientific procedures. It is an objective, impartial, empirical and logical analysis and recording of controlled observations that may lead to the development of generalizations, principles or theories, resulting in to some extent in prediction and control of events that may be consequences or causes of specific phenomena. Research is scientific and as such is not satisfied with isolated facts, but seeks to integrate and systematize its findings. It is concerned with the objective verification of generalizations. Such a verification requires logical analyses of problems and devising of appropriate methodologies for obtaining evidence.

Best (1977, p. 9) has summerized the main characteristics of research as under:

1. Research is directed towards the solution of a problem. It may attempt to answer a question or to determine the relation between two or more variables.
2. Research emphasizes the development of generalizations, principles, or theories that will be helpful in predicting future occurrences. Research usually goes beyond the specific objects, groups or situations investigated and infers characteristics of a target population from the sample observed. Research is more than information retrieval, the simple gathering of information. Although many school research departments gather and tabulate statistical information that may be useful in decision-making, these activities are not properly termed research.
3. Research is based upon observable experience or empirical evidence. Certain interesting questions do not lend themselves to research procedures because they cannot be observed. Research rejects revelation and dogma as methods of establishing knowledge and accepts only what can be verified by observation.

4. Research demands accurate observation and description. The researcher uses quantitative, numerical measuring devices, the most precise means of description. The researcher selects or devises valid data gathering instruments or procedures and employs appropriate mechanical, electronic, or psychometric devices to refine human observation, recording, computation, and analysis of data.
5. Research involves gathering new data from primary or first-hand sources or using existing data for a new purpose. Teachers frequently assign a so-called research project that involves writing a paper dealing with the life of a prominent person. The students are expected to read a number of encyclopaedias, books, or periodical references, and synthesize the information in a written report. This is not research, for the data are not new. Merely reorganizing or restating what is already known and has already been written, valuable as it may be as a learning experience, is not research. It adds nothing to what is known.
6. Although research activity may at times be somewhat random and unsystematic, it is more often characterized by carefully designed procedures, always applying rigorous analysis. Although trial and error are often involved, research is rarely a blind, shotgun investigation—trying something to see what happens.
7. Research requires expertise. The researcher knows what is already known about the problem and how others have investigated it. He has searched the related literature carefully. He is also thoroughly grounded in the terminology, the concepts, and the technical skill necessary to understand and analyse the data that he gathers.
8. Research strives to be objective and logical, applying every possible test to validate the procedures employed, the data collected, and the conclusions reached. The researcher attempts to eliminate personal bias. There is no attempt to persuade or to prove an emotionally held conviction. The emphasis is on testing rather than on proving the hypothesis. Although absolute objectivity is as elusive as pure righteousness, the researcher tries to suppress bias and emotion in his analysis.
9. Research is characterized by patient and unhurried activity. It is rarely spectacular and the researcher must accept disappointment and discouragement as he pursues the answers to difficult questions.
10. Research is carefully recorded and reported. Each important term is defined, limiting factors are recognized, procedures are described in detail, references are carefully documented, results are objectively recorded, and conclusions are presented with scholarly cautions and restraint. The written report and accompanying data are made available to the scrutiny of associates or other scholars. Any competent scholar will have the information necessary to analyze, evaluate, and even replicate the study.
11. Research sometimes requires courage. The history of science reveals that many important discoveries were made in spite of the opposition of political and religious authorities. The Polish scientist Copernicus (1473-1543) was condemned by church authorities when he announced his conclusion concerning the nature of the solar system. His theory that the sun, not the earth was centre of the solar system in direct conflict with the older Ptolemaic theory, angered supporters of prevailing religious dogma, who viewed his theory as a denial of the story of cereation as described in the book of

> Genesis. Modern researchers in such fields as genetics, sexual behaviour, and even business practices have aroused violent criticism from those whose personal convictions, experiences, or observations were in conflict with some of the research conclusions.

From the examination of the above characteristics it is evident that the knowledge gained by research is of highest order. It is not based on assumptions, beliefs, and untested generalizations. To seek such knowledge, a researcher must develop scholarship, accurate observation, integrity, willingness to spend long hours to collect and study all forms of evidence before arriving at conclusions, scholarship and ability to recognize causal relationships, and originality and objectivity in thinking. However, it must be recognized that such ideals serve as goals for which researchers should strive. Being human beings, it is difficult for researchers to realize all ideals completely.

THE NATURE OF EDUCATIONAL RESEARCH

Educational research refers to a systematic attempt to gain a better understanding of the educational process, generally with a view to improving its efficiency. It is an application of scientific method to the study of educational problems. Travers (1958, p. 5) states:

> Educational research...............represents an activity directed towards the development of an organized body of scientific knowledge about the events with which educators are concerned. Of central importance are the behaviour patterns of pupils, and particularly those to be learned through the educational process. A scientific body of knowledge about education should enable the educators to determine just what teaching and other learning conditions to provide in order to produce desired aspects of learned behaviour among young people who attend school.

Since education is a behavioural science, the major concern of educational research is to understand, explain, and to some degree predict and control human behaviour. It is an activity directed towards the development of organized and useful body of scientific knowledge about the events with which educators are concerned.

Needs for Research in Education

Research in education as in the other fields is essential for providing useful and dependable knowledge through which the process of education can be made more effective. There are various considerations which emphasize need for research in education.

1. Education has strong roots in the fields like philosophy, history,

economics, psychology, and sociology. It is, however, based on a conceptual frame of theory. It is through an intensive process of scientific enquiry about the philosophical, historical, economic, psychological, and sociological impact on various aspects of education that sound theories can be established.

2. Education is considered as much a science as an art. As a science, it has a corpus of knowledge concerning the nature of human mind, its growth and development; theories of administration and supervision; educational programmes and practices prevalent in different countries and their results. The quantum of knowledge is indicated by the courses in education prescribed by various organizations and institutions for earning degrees in this field and also by the research material which is being produced and continuously reported in different educational research journals. Since education depends on a corpus of knowledge there is need to add scientific knowledge to it for enrichment and improvement. This will facilitate in making adjustments in educational programmes accordingly. As an art, education seeks to impart knowledge effectively. For example, "How the teacher can play an effective role in the classroom and outside?" is a vital question before educationists, and needs careful research efforts to enhance teacher's effectiveness.

3. The slogan of democratization of education since 1870 resulted in the expansion of education. It has given rise to numerous problems like the problems of individual differences, expansion, buildings, discipline, and so on. Solutions of such problems by trial and error or by experience from tradition and authority often yielded erroneous results. Moreover, growth by experience is very slow and accidental. We need solutions based on research so that the coming generation is not left to the mercy of errors of outright sins of tradition, ignorance and prejudice.

4. There is need for educational research because of the changing conception of education. The International Commission on the Development of Education in its report "Learning to Be" (Unesco, 1972, p. 143) emphasizes:

> Education from now on can no longer be defined in relation to a fixed content which has to be assimilated, but must be conceived of as a process in the human being, who thereby learns to express himself, to communicate and to question the world, through his various experiences, and increasingly—all the time—to fulfil himself. It has strong roots, not only in economics and sociology, but also in findings from psychological research which indicate that man is an unfinished being and can only fulfil himself through constant learning. If this is so, then education takes place at all ages of life, in all situations

and circumstances of existence. It returns to its true nature, which is to be total and life long, and transcends the limits of institutions, programmes and methods imposed on it down the centuries.

In the context of the above nature of education, the limits of educational research have to be extended from the formal and conventional modes of education to the non-formal and innovative systems based on ecological and cybernatic models.

5. During the last two decades great changes have taken place as a result of the rapid scientific and technological developments. Education has to play an important role so that we can accept the change in a smooth way. It can do so by bringing improvements in the existing curriculum, textbooks, methods of teaching, and evaluation.

Historical Development of Educational Research

The scientific era in the physical sciences began in the seventeenth century. The emergence of education as a science is only eighty years old. It was due not only to the complex nature of the phenomena to be studied, but also to slow progress in the development of the measuring tools for collecting data of educational interest. The history of various developments in educational research, therefore, is closely linked with some of the significant developments in the history of measurement and testing.

Educational research as a discipline has passed through several stages before it could take the shape of modern movement for the objective study of educational problems. The history of its growth may conveniently be divided into four phases.

Educational Research Before 1900

As early as 1803 Pestalozzi proposed the scientific study of pedagogy and established his pedagogical school; Sheldon began experimentation at Oswego, New York in 1860. In 1894 Hall printed the first series of questionnaires to be used in securing information about children. His questionnaires mainly dealt with educational topics, such as "The Beginnings of Reading and Writing", "School Statistics", "Number and Mathematics", "Examination and Recitations" and "Some Characteristics and Tendencies of School Children" (Harris, 1960, p. 1162).

Educational research before 1900 was greatly influenced by the experiments in the field of psychology. A number of studies of the threshold of sensitivity were conducted by Weber and Fechner. Wundt started his experiments as early as 1861 and established his experimental laboratory at Leipzig in Germany in 1879. Ebbinghaus conducted his experiments on memory and demonstrated that memory could be measured by some techniques accurately. The famous works of Sir Francis Galton, such as "Hereditary Genius: An Inquiry into Laws and Consequences", published in 1869 and "Inquiries into Human Faculty and its Development", published in 1883 are the evidence of his particular interest in individual differences and mental inheritance. James McKeen Cattell, influenced by Galton, began a systematic investigation of individual differences in reaction time in sensory-motor functions in relation to human intelligence. In 1890, he wrote a classic article called "Mental Tests and Measurement" in which the "Mental Test," was first introduced into the literature. Joseph M. Rice proposed, in 1897, that uniform spelling tests be given to students. He is recognized as the pioneer in the educational research movement.

Educational Research from 1900 to 1920

Early in the century there was widespread interest in the exploration and development of measuring instruments needed by the researchers in the field of education and psychology. In 1904, Thorndike influenced by James McKeen Cattell and others, published the first book dealing directly with the subject of mental measurement. A year later Binet, who had been working with psychological tests in France for about 10 to 15 years, published, in collaboration with Simon, the Binet-Simon General Intelligence Scale. This was followed by the Stanford Revision by Terman. In 1908, Stone, student of Throndike, published his Arithmetic Test, which is considered to be the first standardized achievement test. This was followed by a very large number of tests in several subjects. The Hillegas Composition Scale and Ayres' Hand-writing Scale appeared in 1912. Buckingham Spelling Scale, Starch's Reading Tests, Kelley Kansas Silent Reading Test, the Trabue Completion Test Language Scales, the Thorndike-McCall Reading Scale for Understanding of Sentences, and Monroe's Standarized Silent Reading Tests appeared in the years immediately following the year 1912 (Harris, 1960, p. 1163).

Group tests of intelligence were first developed during World War I, Aptitude tests, such as the Seashore Test of Musical Talent, also emerged during this period.

In addition to the emphasis upon the development of measuring tools, the activities like school surveys were also taking place to promote the scientific study of education during the period 1900 to 1920. The Pittsburgh Survey in 1907 was the first school survey. The

New York City Survey in 1911-12 was the first survey in which arithmetic achievement tests were used. The Cleveland Survey of 1915-16 studied many phases of the school's programme such as student achievement, the curriculum and methods of teaching, administrative organization, school finance, the school plant, staff and personnel, and pupil accounting.

The organizations like the National Society for the Study of Education, the American Educational Research Association, and the National Society of College Teachers of Education were all organized in the USA during the years 1900 to 1920. Many books, monographs and journals dealing with the subject of psychological and educational measurements, statistics, child development, and educational psychology were also published during this period. The Journal of Educational Psychology was founded in 1910 and the Journal of Educational Research in 1920.

Educational Research From 1920 To 1945

The period from 1920 to 1945 is considered a period of self criticism and improvement in educational research techniques and designs. Harris (1960, p. 1164) points out that the four kinds of criticism that greatly influenced educational research methods during this period were: (*a*) research workers gave too little attention to the less tangible aspects of education ; (*b*) research workers provided inadequate theoretical orientation for their efforts; (*c*) single variable laboratory research was inadequate for the complex field conditions actually met in education; and (*d*) educational research gave too little attention to the individual. The research workers in order to be objective concentrated their efforts mainly upon the things that could be readily counted or measured. Among the early counting investigations were the age-grade progress studies, the questionnaire surveys, and the frequency studies. Much attention was given to the measurement of attitudes, thinking process, and personal adjustment with the help of tests and scales. A great emphasis was laid on the concept of validity and reliability of measuring tools during the years 1920 to 1945. The American Educational Research Associations (AERA) produced the Encyclopaedia of Educational Research, the first edition of which appeared in 1940. It gives a detailed account of the earlier research efforts.

Educational Research After 1945

Since 1945, research methods and procedures have been improved with the aim of attaining more objective and reliable information. Many new methods of data collection and analysis were devised. New developments in research design, which were closely related to developments in satistical procedures such as the analysis of variance and of co-variance, made possible field experiments or trials for which randomization provides controls on extraneous variables and for which more than one variable of interest may be studied at the same time. These designs clearly extend the possibilities for the study of complex phenomena in field of education (Harris, 1960, p. 1165).

The scope and fields of educational research have greatly expanded. A number of researches are being conducted all over the world for determining the effectiveness of different aspects of the curriculum, teaching methods, evaluation, guidance, administrative practices and planning. The development of the concept of action research by Corey (1953) has been very useful in improving school practices and in solving educational problems of local nature.

In India the importance of educational research has been recognized rather late. Most of the research in the field of education was conducted after independence. More and more universities started M. Ed M.A. and Ph. D. courses in Education. The organizations like National Council of Educational Research and Training, State Institutes of Education, National Institute for Educational Planners and Administrators, etc. were founded in the country to promote and coordinate educational research at different levels of instruction. Many journals like Indian Educational Review, Indian Journal of Education, Journal of Higher Education, Journal of Education and Psychology, Indian Journal of Applied Psychology, Journal of Psychological Researches and other periodicals were also started to publish information and articles relating to research findings in the field of education in the Indian context. The University Grants Commission, The Union Ministry of Education, the Indian Council of Social Sciences Research, the National of Educational Research and Trainning, and other organizations are providing financial aids to research workers in education and other related fields to undertake and accomplish useful research.

This brief outline has presented only some highlights in the development of educational research in the world. For a fuller and exhaustive discussion on the history of educational research in India and abroad, the reader is referred to the Encyclopaedia of Educational Research (Monroe, 1950; Harris, 1960; Ebel, 1969), The Third Indian Year Book of Education: Educational Research (Adaval, 1968), and First and Second Surveys of Research in Education (Buch, 1973, 1979).

Limitations of Educational Research

In spite of the application of scientific method and refinement of research techniques, tools and designs, education as other social and behavioural sciences has not attained the perfection and scientific status of physical sciences.

It has not been able to establish generalizations equivalent to the theories of physical sciences in the scope of explanation, prediction, and control. Frequently there is lack of unanimity among researchers in the field as what the established facts are or what explanations are satisfactory for the assumed facts. Ary *et al.* (1972. pp. 18-20) have pointed out several limitations which are involved in the application of scientific method in education. They are of the opinion that it is because of these limitations that education like other social scieces will never realize the objectives of science as completely as the physical sciences.

1. *Complexity of subject matter.* The researcher in the field of education deals with the complex nature of human subjects. He is not concerned with a limited number of variables as is the case with a natural scientist who deals with physical or biological phenomena. To understand the complex nature of human subjects he has to deal with a number of variables, acting independently and in interaction. Each individual is unique in the way he grows, in his mental, social and emotional behaviour, and in his total personality. The researcher in the field of education, therefore, has to study the individual in groups and the impact of the behaviour of group members on an individual.

2. *Difficulties in observation.* Observation of human behaviour is more subjective than the observation of physical or biological phenomena. The subjectivity on the part of the observer has a direct impact on the interpretation and findings on which he bases his conclusions.

3. *Difficulties in replication.* The physicist can find objectively and accurately resistence of a wire in the laboratory. He can report his findings and the experiment can be easily replicated by others. This is not easily possible in education. A teacher in India cannot reproduce the conditions of an American teacher's experimental teaching method with the same precision of replication as that with which an Indian physicist can reproduce an American physicist's experiment. Educational phenomena are singular events an cannot be replicated for purposes of observation with same degree of precision and objectivity.

4. *Interaction of observer and subjects.* The researcher in the field of education is a human being. His presence as an observer in a situation may change the behaviour of his human subjects. For example,

the school children may not exhibit their natural behaviour when they know that they are being observed for a particular type of behaviour.

5. *Difficulties in control.* The rigid control of experimental conditions is possible in the laboratory of physical or biological sciences. It is not possible in educational studies in which the researcher has to deal with the human subjects and many variables simultaneously.

6. *Measurement problems.* The tools used for measurement and data collection in the educational studies are much less valid and reliable than the tools of the natural scientist.

Steps in Educational Research

As mentioned earlier that educational research is the application of scientific method to the study of educational problems. The steps in educational research, therefore, are more or less identical to those of scientific method. Only a brief discussion of the steps is presented here, because many of the points are treated at greater length in other chapters.

1. *The research problem.* Educational research starts with the selection of a problem that the researcher identifies from the area or field of his interest. It must be a question that can be answered through scientific inquiry. The problem should be such that can be clearly and precisely stated. The statement of the problem must be complete. It must be presented in a form that makes absolutely clear what data or evidence must be obtained in order to solve the problem.

2. *Formulation of hypotheses.* Educational research should make use of carefully formulated hyhothese. These may be formally stated or implied. In formulating hypotheses, the researcher should keep in mind that the hypotheses are tentative generalizations about the nature of the difficulty under consideration, calling attention to fundamental relationships or possible solutions. The manner of formulating hypotheses is an important aspect of educational research and the researcher should give much thought to it.

3. *The method to be used.* The selection of research method to be used is of utmost importance in the research process. It refers to the general strategy followed in collecting and analyzing the data necessary for solving the problem. The research methods are generally classified in three categories: (*i*) Historical; (*ii*) Descriptive; and (*iii*) Experimental. The method used in the study is dictated by the nature of the problem and the type of data required for answering the questions relating to the problem.

4. *Data collection.* Whereas the research method describes the overall approach to the problem, this step is concerned with the pro-

cedures and techniques to be adopted for data collection. It refers to the nature of the sample to be chosen for study, and selection and development of data gathering devices such as tests, questionnaires, rating scales, interviews, observations, checklists, and the like.

5. *Analysis and interpretation of data.* Good research is characterized by the care taken in the analysis and interpretation of data. It includes the selection of appopriate statistical and other techniques to be used for processing the data collected for the study.

6. *Reporting the results.* This is the last and important step of the research process. It is characterized by carefully formulated inferences, conclusions, or generalizations. The researcher must be able to report his procedures, findings, and conclusions in an intelligent form to others who may be interested in his study and its results.

BASIC AND APPLIED RESEARCH

Generally, research is classified as either *basic* research or *applied* research. This classification is based on the objectives or purposes the researcher intends to accomplish.

The terms basic and applied refer to the extent to which the research is directed towards the solution of educational problems. According to Travers (1948, p. 4):

> Basic research is designed to add to an organized body of scientific knowledge and does not necessarily produce results of immediate practical value. Applied research is undertaken to solve an immediate practical problem and the goal of adding to scientific knowledge is secondary.

Basic research is primarily concerned with the formulation of a theory or a contribution to the existing body of knowledge. Its major aim is to obtain and use the empirical data to formulate, expand or evaluate theory. Some have termed this research as *pure* or *fundamental.* This research draws its pattern and spirit from the physical sciences. It represents a rigorous and structured type of analysis. It employs cereful sampling procedures in order to extend the findings beyond the group or situation and thus develops theories by discovering proved generalizations or principles.

The main aim of basic research is the discovery of knowledge solely for the sake of knowledge. It has little concern for the application of the finding or social usefulness of the findings.

Applied research, on the other hand, is directed towards the solution of immediate, specific, and practical problems. It is performed in

370.78
KOU

relation to actual problems and under the conditions in which they are found in practice. The goal of applied research in terms of adding scientific knowledge acquires only a secondary position. It places importance on a problem here and now.

The applied research also uses the scientific method of inquiry. Its methodology, however, is not as rigorous as that of basic research. Moreover, its findings are to be evaluated in terms of local applicability and not in terms of universal validity. Applied research is mainly intended to improve school practices and add to greater teacher effectiveness in a practical manner. Most of the problems faced by teachers, policy planners, and administrators are solved through applied researches.

Applied research is also designated as *action* research. Some people are inclined to differentiate action research from applied research in as much as the former has relevance only to classroom situations while the latter can probe into practical problems of greater complexity and wider applicability. Action reseach, according to Corey (1953, p. 145). the originator of the term, is the research undertaken by practitioners so that they may improve their practices. Lehmann and Mehrens (1971, p. 6) point out that:

> Action research is a type of applied or decision oriented research, but with the stipulation that the researcher is the same person as the practitioner who will make and live with the decision.

Action research, therefore, is applied research conducted by the teachers, supervisors, and school administrators themselves for dealing with classroom problems and improving classroom practices.

An example given by Turney and Robb (1971, p. 7) would make the distinction between basic and applied research more clear.

> If a researcher were to conduct a study using white rats to determine the effect of positive versus negative reinforcement on learning to run a maze successfully, he would be conducting basic research. If he rewarded one group of rats with food pellets for correct turns in maze, but gave an electric shock to the group of rats making incorrect turns, he could then determine which group mastered the maze in a shorter period of time Even though this information has no immediate application to motivation, it might be useful in the building of a theory concerning motivation and learning.
>
> Whereas food and electrical shock might be impractical to use in the classroom, the researcher could conduct an experiment using praise and criticism as motivational factors in human learning. This type of experiment would be directed towards the solution of a practical problem and, thus, would be classified as applied research.

S.C.E.R.T., West Bengal
Date 29.1.91
Acc. No 4849
S.C.E.R.T., West Bengal
Library

Corey (1953) is of the view that if classroom teachers are to make an active research contribution, it will probably be in the area of applied or action research. Studies can be undertaken for the purpose of improving local school practices, and for teacher growth, professional improvement, and the development of a better curriculum and preparation of good textbooks. Such studies advocate that research should be the function of a group of teachers, with research specialists serving either as consultants or as members of the research team. They provide sufficient flexibility to permit modification of the hypotheses and procedures as the study goes on. Their purpose is not only to improve school practices but at the same time, to improve those who are to improve the practices.

There is no conflict between basic research and applied research. The difference is in emphasis, not in the method or spirit. Each type is committed to the high standards of scientific objectivity and scholarship. The researcher in each type of research attempts to define the problem being studied with precision, to derive his hypotheses from a rich background of information related to the problem, to design the study so that it will result in a genuine test of hypotheses, to collect and analyse facts or evidence carefully, and to draw generalizations objectively. The conditions under which the two researcher carry out their inquiries however, may differ in terms of the control and precision. In basic research, the researcher tries to control a situation by the use of laboratory techniques or other techniques, but the precision is gained at the expense of relevance of the findings. The researchers engaged in applied research conduct inquiries in the complicated psycho-sociological climate of on-going school activities. Because of the multiplicity of variables involved, the research is often lacking in control and precision. The findings, however, have relevance for practice because they derive from a research conducted in a real situation. Moreover, the researcher gets highly motivated because he wants immediate solution of the problem. On the other hand, the motivation for the researeher in the basic research is the intrinsic value of the research and its possible contribution to the advancement of knowledge. Such type of motivation may not have sustaining value for the researcher as it has in the applied research.

It should be noted that certain applications are made from the theory based on basic research. For example, the teacher applies the theories of learning or motivation in the actual classroom situations. The teachers as professional practitioners should be familiar with the findings of basic research, particularly in the areas of their specialization. However, the basic research may depend upon the findings of applied research for building and testing theories.

ABOUT THE POSTTEST

Now you have read the first chapter, you will want to know whether the learning objectives listed at the beginning of the chapter have in fact been successfully realized. For this reason, you are required to answer all the questions in posttest. Similar tests have been constructed for all other chapters.

POSTTEST

1. List and describe various methods by which man from earlier time sought answers to his problems.
2. What is syllogism? Construct two syllogisms: (*a*) one that is logically sound and (*b*) one that is faulty; indicate the nature of the fallacy.
3. What is inductive reasoning? Give two examples of inductive reasoning: (*a*) one that is sound, and (*b*) one that it faulty; indicate the nature of the fallacy.
4. What is scientific method? List and describe its various steps.
5. What is science? Discuss the role of scientific theories in the development of science.
6. List and describe the characteristics of a sound theory.
7. What is research? Describe the main characteristics of research.
8. What is educational research? What are its limitations?
9. Trace the history of educational research since the year 1803.
10. Describe the need for research in education.
11. How has research in the basic disciplines such as psychology, philosophy and sociology, etc. benefited educational practices?
12. List and describe the steps in educational research.
13. Define "basic research", "applied research", and "action research". Give examples to bring out the distinctions between them.
14. Differentiate between "basic" and "action" research. Describe the distinctive features of action research and bring out its need and utility in education.
15. How does research help in the development of educational theory and improvements of classroom practices?

2. Areas of Educational Research

This chapter has the following general purposes. First, it lists and discusses research problems relating to the following content areas of education: educational psychology, philosophy of education, sociology of eduction economics of education, educational adminstration, comparative education, educational measurement and test development, curriculum construction and textbooks. teacher education and teaching behaviour, guidance and counselling, and educational technology. Second, it lists and discusses the research problems relating to the following stages and sectors of education: pre-primary education, primary education, secondary education, higher education, agricultural and technical education, and non-formed education. Third, it lists and discusses the research problems relating to teaching process: instructional objectives, entering pupil behaviour, instructional procedures, and evaluation. Fourth it deals with some priority areas of educational research in India.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. List and discuss research problems relating to the following content areas of education: eductional psychology, philosophy of education, sociology of education, economics of education, educational administration, comparative education, educational measurement and test development, curriculum construction and textbooks, teacher education and teaching behaviour, guidance and counselling, and educational technology.
2. List and discuss research problems relating to the following stages and sectors of education: pre-primary education, primary education, secondary education, higher education, agricultural and technical education, and non-formed education.
3. List and discuss research problems relating to teaching process: instructional objectives, entering pupil behaviour, instructional procedures, and evaluation.
4. Describe some priority areas of educational research in India.

The selection of a problem suitable for research requires the researcher to take the stock of his scholarship, talents and abilities and then to identify the area of educational research in which he is able to make the most successful contribution. The researcher should have an intensive study of the area in which he plans to undertake research.

Before discussing the task involved in the selection of the problem, it is necessary to classify the areas of educational research. The classification of the areas can be made in a variety of ways depending upon how we define *education* and how we interpret the need for conducting research in education.

PROBLEMS RELATING TO CONTENT OF EDUCATION

Education as a process takes into account both the science of education and the art of education. The science of education comprises the corpus of knowledge which is largely responsible for making the art of education more effective. The art of education has relevance to classroom practices and for effective dialogue between the teacher and the pupils. Since knowledge is expanding rapidly in all the disciplines including education, there is need to extend the frontiers of knowledge in areas which constitute the science of education.

Considering this point of view. the fields of educational research can be classified in terms of the following content areas:

1. Educational Psychology.
2. Philosophy of Education.
3. Sociology of Education.
4. Economics of Education.
5. Educational Administration.
6. Comparative Education.
7. Educational Measurement and Test Development.
8. Curriculum Construction and Textbooks.
9. Teacher Education and Teaching Behaviour.
10. Guidance and Counselling.
11. Educational Technology.

Educational Psychology

Research in educational psychology has great significance for a teacher. He has to understand the child he is teaching, understand the teaching-learning process, and constantly strive to improve his method of teaching. The usefulness of various theories of learning for designing conditions that produce effective learning in schools have been the central theme of researches in the recent years. Although much of learning theory emerges from laboratary research, researches need to be conducted under conditions more similar to those existing in classroom. Such studies need not be undertaken in actual classrooms but may involve groups of children in classroomlike situations. Conditions conducive to effective learning, factors helpful in promoting memory and concept formation need attention by the researchers.

Cognitive non-cognitive factors, such as, intelligence, aptitudes, creativity, attitudes, interests, motivation, personality traits, needs and adjustment of pupils provide promising fields of research in educational psychology. Varrious influences of home, neighbourhood, peer relationships, and other social relationships that affect child development, growth, and learning are worth investigating.

Research in educational psychology in India is of fairly recent origin. An analysis of the material presented in the Directory of Behavioural Science Research in India (Pareek, 1967) shows that the research conducted in the various filds of educational psychology in the forty years from 1925 to 1965 is hardly impressive. A bulk of work mostly relates to the areas of ability and aptitudes, personality and social dynamics including attitudes, interests and the like. More than 70 per cent of these studies are M.Ed. and M.A. theses and all such works seem to be mere exercises in research methodology. If one judges the value of all such research efforts one is bound to conclude that the work done in different areas of educational psychology is very meagre both quantitatively and qualitatively and is nowhere near a satisfactory condition. It is, therefore, essential to indicate some of the areas in which a good deal of work maybe done to improve upon the present situation.

One of the most urgent tasks is to study problems, arising within and outside the school, which affect the effectiveness of the teacher in his task of improving the learning of students. The fundamental processes of perception, learning and motivation and their application to the classroom situation need to be studied. Researches in the field of achievement motivation are necessary for social mobility, social justice, and economic progress. Kuppuswamy (1968) while discussing research needs in educational psychology points out that such studies may help to offer suggestions regarding the methods of child-upbringing which might be more appropriate to the new trends in social and economic progress in the country. He is of the view that we need studies in creative thinking so that the teacher and students are helped to play their role adequately and effectively in social and economic progress.

Another important area where a good deal of research can be undertaken is that among the exceptional children. It includes the study of all children who are either handicapped in some way or are gifted. The handicapped children may be physically handicapped, the mentally handicapped or the socially handicapped. Among the physically handicapped are the blind, the deaf and the dumb, or children with defective vision, hearing and speech, delicate children, children suffering from polio and cerebral palsy as well as spastic and paralytic cases. The children of low mental ability having IQ below 75 fall in the category of mentally handicapped children. These include hydrocephalic, microcephalic and the idiots (IQ ranging from 0 to 25). imbeciles (IQ ranging from 25 to 50), and morons (IQ ranging from 50 to 75). Among the socially handicapped children are the destitutes and the delinquents. The gifted children are fast learners and exceptionally bright and have very high general mental ability or special abilities. Shanker (1968, p. 175) points out that study of all these types of children with a view to finding ways and means of their proper location, care, rehabilitation, education and guidance is essential both from the individual and the social points of view.

In case of physically handicapped children, Shanker (1968) suggests that the extent of the incidence of physical handicaps, defects and deformities and the types thereof have to be ascertained. This could be done by undertaking research studies in each region or state to survey in a scientific manner the cases suffering from such physical defects. Then there is need for research to find out ways and means of helping as many children of this type as possible.

With regard to the mentally handicapped children also, there is need to determine the number of children of various degrees of retardation in the various states. There is also need to ascertain the possible causes of mental retardation in children, whether it is caused by poor diet, unhealthy climatic conditions, ill-health of parents, or any other factors. After ascertaining the causes, we need to develop specific methods and techniques for handling the various categories of mentally retarded children so that proper educational programmes through concrete and practical media can be planned and organized.

Similarly, there is need to study the problems of socially handicapped or the delinquent children. According to Shanker (1968, p. 179):

> Since the foundations of delinquency are laid in the home and delinquency is often described as the "home industry", research should ascertain as to how far the joint family system and the parental authority have any influence on this problem.

Apart from this, it is to be known which of the various factors of maltreatment, poverty, negligence, broken homes, bad company, exploitation, maladjustment in the school, and so on are more important in various states and regions in the country, for necessary guidance. Problems in the rural and the urban population could also be compared by means of research.

Research may be conducted in job information, careers and openings available for juvenile offenders, so that they could be suitably rehabilitated in the society to which they belong. For the education, training and upbringing of such children we need teachers who have certain abilities, interest and personality traits, and therefore research is needed for finding out the means for selecting and training such teachers.

A systematic and properly designed effort based on research is needed for the early identification of the gifted children so that they are educated properly to grow at their own optimal speed. It is also to be ascertained as to whether we require separate institutions for the gifted or they need to be educated with the normals so that a separate class of social misfits and intellectual snobs is not created.

A country like India cannot afford to lose creative genius. Thus research in the area of creativity needs immediate attention of Indian researchers. Developing tests for identifying creative children deserve top priority.

To help the curriculum constructors and textbook writers, research is needed to have a through understanding about the nature of concept formation and development among the Indian children. Studies (Aaron, 1968) show that time concept, concepts of space and area to not reach a mature level until the age of 11.

In India, Education Psychology is to be recognized as one of the most important fields of educational research. This is the first and essential requisite for the emergedce of good research in education. Deo (1968) too is of the opinion that educational psychology deserves status as an important field of educational research. It failed to get due recognition for the simple reason that those who call themselves educationists are scared and reluctant to admit educational psychology to the realm of education for fear of losing their own importance.

Philosophy of Education

Education has been called the dynamic side of philosophy or an active aspect of philosophic belief. Its theory and practice can never attain perfection unless it is based on the sound footing of a systematic philosophy.

Of the various philosophical dimensions, logical dimensions reveal various significant roles which logic can perform in the formulation of educational concepts, generalizations and analogies, and in drawing out of educational inferences. Logic can also help in curriculum framing and it can provide the very basis of the teaching-learning process. It is, therefore, needed to find out as to how many possible logical operations are involved in teaching-learning process.

Ethical and epistemological dimensions help in formulating the theory of education in its most general phases. In examining the aims of education, the motivation of learning, or the measurements of its results, we are dealing with ethical problems, the problem of values. Ethical considerations also come up unavoidably in examining the social or political setting of the educative process. Values are also important considerations in selecting which studies should be included in the curriculum. The curriculum also raises some very important questions of epistemology. The curriculum being the student's avenue of approach to knowledge, it makes us understand the nature of knowledge. The nature of knowledge not only have an influence on the way in which the curriculum is organized and taught, but it also helps to determine the shape of teaching-learning strategy.

In view of the demand for a philosophical foundation of education, there is an imperative need for undertaking research into the various problems pertaining to its different dimensions.

In India, there is need for the development of a philosophy of education from the Indian stand-point. The aims and means of education are to be deduced from the *Indian* conception of life. A study of the utterances of Vayasa and Valmiki, Buddha, Sankaracharya, Swami Vivekananda, Sri Aurobindo, Tagore and Mahatma Gandhi would help us in formulating our own system of education.

A philosophical analysis of the problems of indiscipline, unrest strikes, disobedience of authority and the like in the context of the prevalent circumstances and conditions would be highly significant and fruitful.

Naidu (1968) has listed following areas of philosophy of education in which research can be undertaken:

1. Need for Philosophy of Education (This is just like pointing out the need for axioms and postulates in geometry).
2. Special features of Indian Philosophy.
3. Indian Philosophy and Indian Education.
4. Regrouping of Indian Philosophic Thought under the following heads:

- (*a*) Materialistic, Physico-mentalistic, Supra-mentalistic.
- (*b*) The self and the non-self: the self in these systems, their inter-relation.
- (*c*) Discussions of the following problems from the stand-point of the three trends treated comparatively and critically:
  - (*i*) Problem of Matter: Maya, Space and Time, Casuality.
  - (*ii*) Problem of Mind: Mind (manas), Body, Ego and Buddhi.
  - (*iii*) Problem of Truth: Knowledge and Truth.
  - (*iv*) Problem of Beauty.
  - (*v*) Problem of Goodness.
  - (*vi*) The Supreme Value of Self-realisation.

5. Philosophy of education deduced from 4. This may be done under three major heads of exposition:
   - (*a*) The Teacher—who is a teacher?
   - (*b*) The Learner—who is competent to learn?
   - (*c*) Teacher-learner Inter-relation and Interaction:
   - (*i*) Aims of education (derived from the great Indian values).
   - (*ii*) The unique method of teacher (Swami Vivekananda).
   - (*iii*) Indian concept of discipline.
   - (*iv*) The educative process and its content (what should be taught taking into consideration modern requirements).
   - (*v*) The outcome of the educative process and how it should be evaluated.
   - (*vi*) The final charge to the outgoing student.
6. Some Ancient Educational Institutions which practised the philosophy outlined above.
7. The abiding elements in Indian Philosophy of Education.
8. Reorganisation of Education to make it Indian on the basis of the exposition attempted so far.

Moreover, there is a problem of value-crisis in India today and it is because of this problem that our entire system of education is drifting aimlessly. In this context, Bhatt (1968) suggests that educators and educational administrators in India should wake up to the task of the reconstruction of national scheme of education and its goals by undertaking a critical and thorough examination of the present set up of values—their nature, status, types, order, spheres, standards, and so on.

Sociology of Education

Sociological factors have a great impact on the educational thought and practice. The study of all such factors is of great practical significance. According to Travers (1969, p. 47):

> Although the central topic of educational research is the development of a science of behaviour in educational situations, peripheral studies pertaining to sociological and anthropological problems are of considerable importance.

Some important problems of research related to sociology of education are the study of population explosion and changes, demographic trends and the impact of political and social pressures on an educational system, Subjects like educational aims, curriculum contents, methods and techniques in teaching-learning process conducive to socio-economic and politico-cultural situations prevailing in the country from time to time deserve attention by researchers.

The role and function of the school should be determined in the light of the needs of the specific community background. In the problem of delinquency, the study of social factors and cultural background of the community are also important. Problems relating to tribal cultures, rural community, community development, industrialization, urbanization, crime and family need to be investigated thoroughly.

In India, very few researches have been conducted in the field of sociology of education. The problem like the teacher's role as an agent of social change and modernization, teacher's morale, the impact of population explosion on childrens' admissions and academic achievement in schools, social determinants of educability of school children, the role of education in family planning, sociology of defferent tribes, and social obstacles in the inculcation of democratic and secular values among school children need to be investigated on the priority basis.

Economics of Education

Economics of education is a rewarding area of educational research The major objective of research endeavour in this area is to apply effectively the relevant economic analysis to the field of education with a view to raise the efficiency of educational system at different levels and to identify the role of education in social and economic development of the country.

Thorp (1962), Harris (1962), Schultz (1963), Becker (1965), and Kershaw and Mckean (1959) have suggested useful areas of research in the economics of education. The lead provided by these researchers can be profitably modified and utilized for studies in Indian conditions.

In the Indian context, Singh (1968) points out that problems of optimum size of an educational institution and resource allocation pose severe challenges because of the formidable obstacles in measuring educational inputs and outputs. Application of marginal and mathematical (linear programming) analysis for the determination of optimum size of our educational institutions and least cost combinations amongst different costed techniques of a given institutional unit are some problems for investigation. Construction of *efficiency indices* of school units would also be helpful. In view of the market complexities the determination of *shadow prices* at conceptual level in respect of teacher salaries and other inputs of the system may provide valuable information. Analysis of cost behaviour of educational institutions is also an important area of research. In this connection identification of cost determinants at various levels will be of special importance while planning for facilities for education. Studies aiming to determine the pattern of educational expenditure at vasious levels, grant-in-aid, capital outlays, unit costs and opportunity costs may also be conducted.

There is a need for economic evaluation of various educational programmes undertaken by the state and central governments. These include agricultural education, vocational education, workers' education, adult education and extension eduction.

Educational Administration

Research in the area of educational administration should aim at expanding knowledge which may help to explain various relationships underlying the process of educational administration and in building a theory of educational administration. It should also help in finding solution to problems related to the administration of education so as to make it efficient and effective. To realize these aims, it would be worthwile to undertake research in the areas of staff personnel administration, educational planning, educational legistation, school plant planning, school organization, business administration, evaluation of institutions, administrative theory and supervision.

The quality of research conducted in the area of educational administration in India is of a poor standard. A review of such researches conducted during the period 1939 to 1980 indicates that most of the studies on educational administration in Indian universities are a part of requirements for M.Ed. and M.A. (Education) degrees. Sharma (1968) has suggested that the studies which will have implications for developing policies in respect of recruitment, development, in India. Job analysis of a headmaster, principal. sub-inspector, inspector, district education officer, deputy director and director in the Indian context would be fruitful studies in the area of educational administration and supervision. It would also be worthwhile to study the supervisory role of a secondary school principal as well as that of a elementary school headmaster. Researches to study the place and scope of collective bargaining in education, the impact of existing compulsory education laws, legal status of voluntary organizations in the administration and control of education also need attention. Techniques to improve the efficiency of school organization need to be explored so as to maximise learning by pupils.

Comparative Education

Political, cultural, socio-economic and geographical conditions have an impact on the educational systems of different countries. Comparative study of educational systems of various countries is a fruitful area of research. Giving the aims of comparative education King (1965) holds that:

> The full value of comparative education as a discipline in its own right is appreciated when we realize the three-fold role it can play in this partnership of communication. Firstly, it can contribute to the very important basic concept of a *culture* as the nucleus of any areas study not just in the narrower anthropological sense but with proper attention paid to consciously formative institutions like any instruction, schools or ways of life . . . . Secondly, it can adduce relevant evidence from comparable situations in other cultures, to be measured against the phenomena being studied in the area under review; for cultural factors may differ from situation to situation though the economic or political factors may seem identical. Thirdly (and without claiming hegemony), it can act, as a kind of ghostly direction-finder or pointer for some of the specialized academic disciplines themselves, which it can compare and evaluate in an intercultural perspective.

These three aims suggest that researches in comparative education need be culture-based. In this context, Singh (1968) suggests that intracultural studies and studies of inter-cultural types should be taken up.

Another area of research in comparative eduction could be of making intensive area studies so as to gain insights into particular systems of education. Analysing the educational organization and administrative machinery of two or more countries is also an useful research area.

Different parts of the world have come closer because of the conquest of space and distance. Due to this fact peculiarities of a particular system of education are fast disappearing and the problems of education are getting common to all countries, The attempted solutions, according to Singh (1968), might differ in their texture but the purpose behind these attempts remains essentially the same. Problems like education and national development in terms of economic growth, problems of educational control and reorganization of curriculum, the role of universities, social education, and so on are a few topics on which purposeful research is possible.

For India, Education Commission (1966) has recommended a two-stream pattern of secondary and higher secondary education. This idea is similar to that of British Fifth Form and the Sixth Form. The Commission has further recommended that specialization should begin at the higher secondary stage as is done in the British Sixth Form. Bifurcation of students into general and vocational streams at the end of lower secondary stage has been borrowed from the Russian system. Before implementing the recommendations of the Commission, it would be desirable to make a thorough study of the British and Russian patterns along with their adaptation in other countries. Likewise, the innovative ideas of *work experience, school complex, setting up of major universities in the country and granting autonomous status to some colleges* and so on need a comparative probe into them.

Comparative research programmes at the national level can also be undertaken to cover all the states and union territories for comparison, while the state level programmes can be confined to the comparative study of one or two other states, preferably the neighbouring states because of their cultural affinity. In this context, Odd (1968) has suggested that inter-state study in grant-in-aid system and administrative patterns would be fruitful. Some states have nationalized the school textbooks, while others have a free market system. It will be worthwhile to make a comparative study of the procedures, problems of textbook production, the quaiity of textbooks produben, and teachers' reaction to them.

Educational Measurement and Test Development

Educational measurement is primarily concerned with the quanticfiation of certain educational outcomes in the form of the acquisition of certain attitudes, behaviours and skills We need valid and reliable tools for measuring such educational outcomes. These include psychological tests, observational techniques, rating scales and other non-testing devices.

Psychological tests are developed and are used primarily for the determination and analysis of individual differences in general intelligence, aptitudes, educational achievement and non-intellectual personality characteristics. Tests of general intelligence and of any specific aptitude have had their extensive use in educational classification, selection, and planning. Closely associated with the testing of intelligence and aptitudes is the measurement of educational achievement and construction of achievement tests. Achievement tests are not designed primarily for prediction, but they are intended to measure the individual's actual learning in educational subject matter after a period of instruction.

Research in educational measurement and test development is concerned with the critical evaluation of the existing forms of tests, the construction and standardization of valid and reliable tests for measuring the educational outcomes of teaching specific subjects and of various educational activities.

Effective evaluation of students' attainment with respect to accepted goals of instruction is considered an indispensable aspect of good teaching. Written examinations and tests have commonly been used as the chief evaluation device and have dominated our educational system to a great extent. These are often considered the sole criterion for judging a student's competence for various activities. But the evils of the examination system in our country are well known to all. The baneful effects of the system on education in general, and secondary and university education in particular, have been discussed in the reports of several committees and commissions. According to the University Education Commission (1948), if any reform in our education system has to be brought about, it must first be brought about in the examination system. Moreover, the whole examination system has become a law-and-order problem rather than a test of learning or knowledge. Nevertheless the importance of the system has been accepted by all. The University Education Commission (1948) and other subsequent commissions have suggested merely the improvement of the examination system but not its total abolition. At the seminar on examinations in higher education organized by the Inter-University Board of India and Ceylon in January 1971 at New Delhi, it was felt that there is need to improve upon the existing system and to streamline it by ensuring greater reliability as well as objectivity. The crucial question, therefore, is not of the abolition of the system but making it a step in the process of teaching and learning. It is from this point of view that a number of reforms have been suggested by various academic and research organization like the UGC, the State Boards of Secondary Education, and Examination Reform Units of various universities. Consequently, a number of significant innovations like grade-system of marking, discontinuation of pass/fail system and declaring results subject-wise, semester-system of examination, orientation of paper-setters and evaluators, improvement of evaluation instructions for the examiners/evaluators, supplementing essay-type tests by objective type tests, supplementing written tests by oral tests, re-evaluation of answer-scripts, supplementing external public examination by periodical internal assessment, spot evaluation, and using mechanical devices for compiling results, have been introduced at the school, college and university levels. It would be worthwhile to study the nature and effectiveness of these innovations at different levels of instruction.

Achievement tests for various subjects at different levels in different languages need to be constructed and standarized. These would be helpful in comparing the achievement of students passing their examinations from different School Education Boards and Universities otherwise the comparison just on the basis of their examination marks becomes difficult.

There is also a great need for developing different types of aptitude tests for evaluating the specific abilities of students. The available aptitude tests are mostly of foreign origin which need to be adapted to the conditions of this country. Besides these specific aptitude tests, different test batteries should be developed for assessing suitability for specific courses of study or jobs. Chatterjee and Mukerjee (1968) are of the opinion that these batteries can be used for estimating success of an individual in different training courses and the marks obtained can be used for identifying unsuitable candidates.

There are several aspects of students' growth that cannot be measured solely by written examinations. Such aspects which include various personality-traits, interests, needs, and values are important for determining success in certain educational and vocational fields such as medicine, social work, teaching and the like. It is, therefore, desirable to develop some tools for measuring such attributes which would require elaborate and carefully planned research work.

The field of educational measurement and testing has rich potentialities for research. Mitra (1968) has suggested the need for a Central Testing Agency, which will act as a coordinating agency having a pool of psychologists, statisticians, items writers, computers and so on. This agency will go a long way to help research on test development, not only by providing others with the necessary facilities, but also by way of bringing together the researchers to think on important problems of test development in India.

Curriculum Construction and Textbooks

The nature and scope of research in curriculum could be easily understood if the meaning of the term 'curriculum' is properly analysed. According to Good's (1959) *Dictionary of Education* the word curriculum means:

1. A systematic group of courses or sequence of subjects required for certification in a major field of study, for example, social studies, physical education curriculum.
2. A general over all plan of the content or specific materials of instruction that the school should offer the student by way of qualifying him for certification or for entrance into a professional or a vocational field.
3. A body of prescribed educative experiences under school supervision, designed to provide an individual with the best possible training and experience to fit him for the society of which he is a part and to qualify him for a trade or profession.

It is evident from the above analysis that the word curriculum is a comprehensive term and it means the organized programmes of learning or tracks of study in the field of knowledge or human experience.

Travers (1969) has pointed out some areas of curriculum research which need consideration by researchers in the field of education. One of these is concerned with the structure of subject matter. Mathematics, for example, has a logical structure and the order in which mathematical concepts are taught should be based on this structure. History also involves many different structures. One of these is time. Similarly, an appreciation of the history of religion involves an understanding of the structures of philosophical ideas. It is worthwhile to evolve simple and well-defined structures in various subjects of study that teaching should involve at particular age and grade levels.

Another area proposed by Travers (1969) is concerned with the development of techniques for making analysis of the psychological demands placed on the learner by various learning experiences. The development of audio visual materials and their impact on the learning of subject matter also falls within the area of curriculum research.

General principles of curriculum building analysis of curriculum in various subjects, organization of curriculum, adaptation of curriculum to local needs, effects of examination on curriculum organization, adaptation of curricula to children's mental abilities and physical development, analysis of textbooks, concept development in various subjects, effectiveness of various teaching methods, human relation in curriculum development, study habits, duration of school work, revision and modernization of curricula are some important aspects of curriculum research mentioned in a bulletin of UNESCO (1951).

In a developing country like ours the school curriculum should change with changes in the society. It has remained static and has little relevance for the actual needs of the youth and society. Education Commission (1966, p. 183) has also pointed out that the school curriculum is generally criticised as being inadequate and outmoded, and not properly designed to meet the needs of modern times. For its upgrading there is a need for systematic curriculum research so that the revision of the curriculum may be worked out as a well coordinated programme of improvement on the basis of the research findings. The preparation of suitable textbooks and other teaching materials is also basic to the success of any attempt at curriculum improvement. Thus not only does the suitability of the existing textbooks need to be verified, but there also seems to be a need for investigation into the best types of textbooks. The research would also help us in solving the problems that relate to the vocabulary, content, printing and illustrations of textbooks.

Teacher Education and Teaching Behaviour

Teacher education is a term rather recent in its origin. It refers to the total educative process which contributes to the preparation of a person for a teaching job in schools. The educative process includes the programme of courses and other experiences offered by teachers' colleges or schools and universities for the purpose of preparing persons for teaching and other educational service and for contributing to their growth in the competency for such service. Thus research into the goals of teacher education and the means of achieving those goals in terms of curriculum, syllabi and programmes is essential for the qualitative improvement of education.

The task of identifying meaningful and measurable criteria of teacher effectiveness is crucial in the field of teacher education. Monroe (1950) and Harris (1960) on the basis of some researches have brought out that there is no unanimity among the researchers about the criteria of teacher success or effectiveness. In view of the lack of agreement, more research is needed in the field of teacher effectiveness so as to provide some generalizations regarding the predictability of teacher effectiveness.

Most of the recent developments in the process of teaching and learning reflect a shift in attitude of researchers with regard to the assessment of teacher effectiveness. Flanders (1969) hopefully sees this as a movement towards more objective approaches. Now it is becoming possible to identify certain teaching activities and say with confidence that these activities have a direct influence on pupil achievement and attitude. The identification of such activities could be very fruitful topics for research. Researches to study the effectiveness of various techniques like interaction analysis, micro-teaching and simulation in modifying the classroom verbal behaviour of teachers could also be undertaken.

Teacher's personality is one of the important variables in the teaching situation, It is the basis of much of the argument that teachers are born and not made. Research will help us in identifying the personality characteristics that influence teaching.

In India, teacher education has been comparatively neglected in post-Independence period though its significance has been stressed by the University Education Commission (1949), the Secondary Education Commission (1953), the International Team on Teachers and Curricula in Secondary Schools (1954), and the Education Commission (1966). It is needed to improve the entire framework of teacher education. For there should be improvement in curricula, programmes evaluation, textbooks and instructional materials for teacher education The curricum suggested by Education Commission (1966) for the professional education of primary and secondary school teachers may be introduced in some teacher training institutions on experimental basis and its impact on the teacher trainees may be studied for comparison. It would also be desirable that the teachers under training should be introduced to a number of new techniques and methods like the use of radio, television, films, programmed texts and language laboratories in classroom situations. The effectiveness of these techniques and methods in developing the professional competence of teacher trainees can be fruitful topics for investigations. Some programmes in the form of summer institutes, refresher courses, seminars and workshops are undertaken for the reorientation of inservice teachers. A follow-up research is needed to study the impact of such programmes on inservice teachers.

Sufficient research has already been undertaken in some foreign countrier in the modification of teaching behaviour through the use of interaction analysis, micro-teaching, team teaching and simulation. There is need of such studies in our country also as the problems and environmental conditions of Indian teachers are different from those of other countries.

Guidance and Counselling

The aims of guidance are both adjustive and developmental. Its function is not merely that of assisting students in making educational and vocational choices but to help them in making the best possible adjustment to the situations in the educational institutions and in the homes. The process of guidance and counselling make use of various methods and techniques.

The field of guidance poses problems of diagnostic and prognostic research. It is closely related to the field of test development. We need to develop adequate tools to diagnose the talents among children at an early stage. The construction of suitable intelligence, aptitude and creativity tests, attitude scales, personality and interest inventories offers worthwhile topics of research.

Research in guidance and counselling should take into account the special conditions and limitations that exist in our country. It should be oriented to our special needs, conditions and culture. According to Harper (1960):

> Unfortunately, too much research done by Indians is also uniquely British or uniquely American—and thereby far less useful than it be to our fellow-countrymen.

He suggests that the joint family and its effects on vocational adjustment of Indian youth; the influence of widely different religious beliefs and attitudes on vocational choice and adjustment; the impact of age-old system of hereditary occupations on the vocational choice of the students; selection of various methods and techniques of guidance and counselling suitable for Indian culture can be very fruitful topics for research.

More research is needed on long-term manpower needs. Exploration of future needs, and probable supply, should be done in all the important fields. The factors which lead to success in various professions and in jobs in industry, business, etc. should be investigated with the help of empirical research. The factors include the study of aptitudes, socio-economic status, attitudes, interest patterns and other personality dimensions. Tests of intelligence, special aptitudes like art, science and music, special skills and scholastic aptitudes need to be developed for Indian students of different age-groups. The important foreign tests which are widely used in different states of India for guidance and counselling purposes should be adapted to India's special conditions.

Educational Technology

The term educational technology has generally come to have two different meanings. One meaning refers to the detailed application of the psychology of learning to practical teaching problems. This is essentially a software approach which is closely associated with the modern principles of programmed learning.

The second meaning of educational technology refers to the application of principles of engineering and techology in the development of electro-mechanical equipment used for instructional purposes. This concept is essentially a hardware approach. It stresses the use of motion pictures, tape recorders, teaching machines, television, projectors, and computers in teaching-learning process.

The development of software and hardware and their impact on the learning of students suggest some important problems for research We need to develop programmed texts for the use of children in India in various subjects for different grades. The suitability of this material in Indian conditions may also be explored. The effects on the academic achievement and other cognitive and non-cognitive factors of the students following programmed instruction material could be studied under Indian conditions and background. The relative benefits of face-to-face instruction and televised instruction in the teaching of different subjects need considerable attention by the researcher in the field of educational technology.

PROBLEMS RELATING TO DIFFERENT STAGES AND SECTORS OF EDUCATION

Education may also be considered as a process embracing various stages and sectors. Longitudinally and horizontally the educational research may be classified in terms of the following areas:

1. Pre-primary Education.
2. Primary Education.
3. Secondary Education.
4. Higher Education.
5. Agricultural and Technical Education.
6. Non-fromal Education.

Pre-primary Education

Pre-primary education is of greatest importance in the child's physical, emotional and intellectual development. Studies have shown that children who have been to pre-primary school show better achievement at the primary stage and help in reducing wastage and stagnation. It is very useful for the children of working mothers and the children coming from slum areas and poor families.

Pre-primary education has been rapidly gaining in popularity in the post independence period. Our advance in this sector, however, has been restricted on account of the inadequacy of the resources available and especially because primary education is being accorded the highest priority. Therefore, ways and means are to be explored so that the progress in this sector may not get retarded.

Some people are of the view that the pre-primary education should be completely left to private enterprise while others prefer that the responsibility for it should be largely assumed by the states. Research in needed to make a comparative study of the two points of view taking into consideration the role played by the private and public sectors in others developed and developing countries.

Research is also needed to settle the competing claims of quality and quantity. Some educationsts are of the view that the pre-primary education to be provided must be of the proper quality, and they suggest that there should be no large scale expansion at the cost of quality. Others advocate the adoption of less costly techniques so that the benefit of pre-primary education may be extended to a large population of children.

The problems relating to the existing pre-primary schools need to be investigated. Research will assist in providing training facilities to the pre-primary teachers and in the preparation of materials and literature needed for pre-primary education. It will also help in providing suitable guidance and supervision to pre-primary schools and training institutions. The problems relating to the objectives, curriculum, and methods of teaching at the pre-primary stage demand immediate attention of the researchers.

Education Commission (1966, p. 150) has suggested that there is need for more coordination among different agencies that work for child care and pre-primary education, both at the national and at the state level. It would, therefore, be worthwhile to determine the roles of the different agencies like CSWB, the Indian Council of Child Welfare and the Community Development Administration, etc. so that an integrated effort can be planned for achieving the objectives set by the Education Commission.

Primory Education

The Directive Principle contained in Articie 45 of our Constitution provides for free and compulsory primary education for all children in the age group of 6-14 years. But in spite of ardous efforts, only about 83 per cent children in the age group 6-11 years and about 37 per cent children in the age group 11-14 years are going to school. These data show that the Constitutional Directive has remained unfulfilled and there is an insistent demand that the centre and the states should undertake every possible effort for hastening the process of universal primary education. Education Commission (1966, p. 149) has also strongly supported the demand that government should work for early fulfilment of the Constitutional Directive because the provision of free and universal education is important "not only on the grounds of social justice and democracy, but also for raising the competence of an average worker and for increasing national productivity." The research in the problems relating to primary education thus assumes special importance.

Unfortunatety, the problems of primary education have received very little attention. This situation has to be changed very early. Naik (1968) has suggested that competent persons should be induced to specialise in problems of primary education and special institutions should be established for conducting research in elementary education. He has indicated some broad areas and problems of primary education in which research is urgently needed.

One of the major objectives of primary education is to entrol every child in the age group 6-14 years into schools. Research is needed to determine the main causes of non-enrolment and to explore the methods of overcoming them. Problems relating to children not attending the schools, the causes that prevent their enrolment in schools, and manner in which an education programme could be organized for them should be explored. The causes of absenteeism of children who are enrolled in the schools have to be studied separately for each individual case. Such studies will provide sufficient data on the nature of absenteeism and its causes and will also suggest remedial measures to meet them.

To eradicate the evils of wastage and stagnation some studies in the field of primary education were conducted in Maharashtra. The findings of these studies have clarified the nature of the problem and evolved some useful concepts for future research. However, according to Naik (1968), no such work has been undertaken in other parts of the country. He is of the opinion that such studies should be repeated from time to time so that the progress made in the reduction of these serious attempt to measure standards has been taken so far. Therefore, it will be worthwhile to study to what extent and in what respects the standards have been falling. Research is also needed to know the gap between the standards in our country and those in the advanced countries.

The unrest among the students in the universities and colleges, the problem of recruitment of university and colleges teachers, the controversy of quality versus quantity, development of library facilities, improvement of teaching, the medium of education, participation of students in decision-making bodies of the universities, the role of student unions, future enrolment policy in higher education, the problem of selective admissions, the role of higher education for women, the development of research facilities, autonomy within the university system, the examination reforms at the university stage, and role of correspondence courses in the higher education suggest worthwhile problems for research in higher education.

Agricultural and Technical Education

Education Commission (1966, pp. 348-388) has recommended that agricultural and technical education should also form a part of general education. All primary schools including those situated in urban areas should give an agricultural orientation to their programmes. Agriculture and technical education should be made an integral part of work experience at primary and secondary school stages. Some agricultural and rural problems should also be introduced in all teacher training programmes. Research, therefore, is needed to set up the objectives of agricultural and technical education in the light of man-power needs. There is an infinite scope for the development of research in devising curricula and evaluation procedures in these fields. Research can also help us in the preparation of textbooks and other instructional materials for imparting agricultural and technical education at different levels of instruction.

Non-Formal Education

Non-formal education includes a variety of programmes such as social education, part-time education, continuing education, orientation and refresher courses, adult-education, extension services, life-long education, balwadis and anganwadis, non-formal education for 6-14 age group, farmers' functional literacy programme, functional literacy for women, non-formal education for workers, distance education, education through mass media etc. The exact definition of non-formal education has become a difficult enterprise because of a variety of educational programmes covered under this area. Since non-formal education programmes are mostly designed to cater to the needs of the adults and children in 6-14 age group, the area may be classified into two broad sub-areas of (i) adult education and (ii) non-formal education for children in the age group 6-14.

The report of the International Commission on the Development of Education (UNESCO, 1972) has provided many possible meanings of adult education. For a very large number of adults in the world today, it is a substitute for the basic education which they have missed. For many individuals who received only a very incomplete education, it is the complement to elementary or professional education. In India, adult education, in fact should serve as a means of individual development. Its scope has to be widened to include *adult literacy* for educating the illiterates and *life-long education* for neo-literates and literates.

Adult literacy is not only a delayed instruction which the adults should have received during their early years but it is the education which would enable the individual to discharge his adult responsibilities meaningfully and fully. Adult literacy programmes, therefore, must go beyond the frontiers of mere literacy programmes so as to be functional in the true sense of the term. It is remarkable that the Working Group on the adult Education for the term plan (1978-83) under the guidance of the Union Ministry of Education and Social Welfare, Government of India, after going through the past experiences, has looked upon *adult literacy* as a method of human resource development which can contribute to new development strategy in three ways. First, by making a substantial part of the work force literate and better skilled; secondly, by the involvement of people in various development programmes which would make it possible to achieve optimum potential and minimize wastage; and thirdly, by the creation of awareness among the poor regarding laws and policies of government whereby it would be possible to implement the strategy of redistributive justice. To achieve these declared objectives a special technology of adult education must be devised with the help of educational research.

For complete eradication of illiteracy, we may have to recognize the importance of mass-cum-selective approach. The selective programme is to be adopted for specific groups of adults, in the age of 15 to 35 years, which will be identified from factories and industries big agricultural farms, mines and other organizations by conducting surveys.

It would be worthwhile to make a well designed survey of adult learners to ascertain their psychological, civic and vocational needs. The knowledge of these needs would be helpful in the preparation of instructional material. The instructional material may include books, newspapers, and periodicals selected on the basis of content, language style, and illustrations.

Adult literacy drive should not relay heavily only on school teachers, but it should include teachers and students of colleges and universities, village youth, retired servicemen and other developmental functionaries. Success in wiping out literacy has been achieved in countries like *Cuba*, where even high school boys and girls were mobilized during the vacation period to teach the adults in the country side. There is no reason why similar experiments cannot be tried in a non-communist country like India. Moreover, it would be worthwhile to conduct studies in motivation of adult learners especially in the rural, backward and far-flung hilly areas of the country. Such studies will also be useful in case of women and persons belonging to scheduled castes and scheduled tribes which constitute 'a substantial and an integral segment of the labour force and are the worst exploited groups in India.

Also, if India is to march ahead on the road to prosperity and keep pace with technological developments in the world, it cannot afford to be oblivious to the need of *life-long education* for the neo-literates and literates. Educational planners are of the opinion that a nation which has not been able to provide the basic elementary education to all the children in the age group of 6-14 years can ill afford to provide for the life-long education of neo-literates and literates. This argument does not seem to hold good as life-long education not only enable individuals in society to cope with the explosion of knowledge, speedy scientific and technological developments all over the world, but also revitalizes the intelligentsia to play a more effective role in regenerating economically and socially handicapped sections of the society.

One of the basic aspects of life-long education is that it does not terminate at the end of formal schooling. It covers all those organized and non-formal means whereby people of all ages are encouraged to learn from one another and from the world around them. These means include correspondence courses; work-study schemes; self-learning or study centres; extension service in education; agriculture and industries; health education services; land reform education programmes; community development education; job training schemes; population education; environmental education; and consumer education. An objective analysis of the suitability of all these means to the prevalent circumstances and conditions of the different parts of the countra would be highly significant and useful. To ensure marked achievement in the goals of adult education, research is needed to evolve a system which would abolish the dichotomy between work and literacy (Koul, 1979).

The non-formal education programmes for children of 6-14 age-group gained momentum only during the last decade when the country accepted non-formal education as an important alternative strategy of education to attain universalisation of education. There is need for undertaking research studies relating to philosophy and goals of non-formal education especially in the context of the children of disadvantaged and deprived sections of the society; population characteristics, needs, and motivational aspects of the childrer participating in non-formal education programmes; development of learning materials, methods and instructional aids; use of mass media; orientation and training of teachers for non-formal education programmes; evaluation of learning outcomes of the clientele; evaluation of non-formal education programmes; and integration of non-formal education with other formal education systems.

PROBLEMS RELATING TO TEACHING PROCESS

Considering education as a purposeful planned activity intended for modifying the behaviour patterns of pupils, we can make use of Glaser's Basic Teaching Model (1962) for classifying the areas of educational research. He has used the following paradigm to explain the teaching process:

This paradigm suggests the following four important areas of educational research:

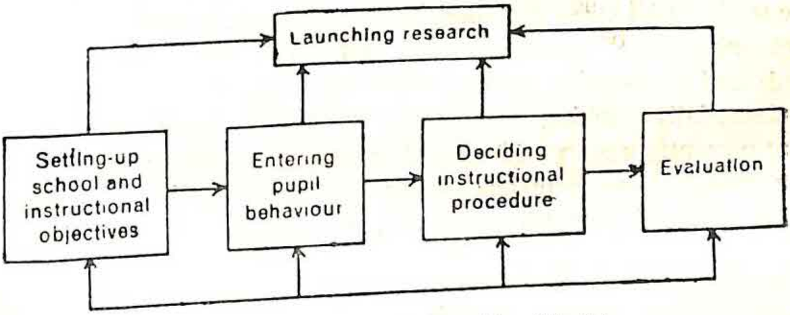

Launching research
Setting-up school and instructional objectives
Entering pupil behaviour
Deciding instructional procedure
Evaluation

FIG. 2.1. A Basic Teaching Model

1. School and Instructional Objectives.
2. Entering Pupil Behaviour.
3. Instructional Procedures.
4. Evaluation.

School and Instructional Objectives

School objectives refer to the broad goals which a nation pursues through its educational system and the administration of its schools. These are broad statements of goals and reflect economic, political, and social philosophies. Objectives which describe a proposed change in the behaviour of a learner as a result of some learning experiences are termed as instructional objectives. School objectives serve as general framwork for stating the narrower and specific instructional objectives.

In the light of the changing economic, political and social factors it becomes necessary to study the school/general objectives. The needs of pupils and society are also to be ascertained in terms of the future projections about man-power needs. The entire educational programme is subordinated by the fixation of school objectives. Besides, the teacher can decide upon the instructional objectives for teaching of each unit of work according to the Taxonomy of Educational Objectives as suggested by Bloom and his associates (1956).

Entering Pupil Behaviour

The recognition to individual differences forms the basis for classroom teaching and therefore, the teacher has to enter into pupil behaviour before he plans the teaching of a fresh unit. The pupil behaviour may be analysed in terms of his achievement, intelligence, aptitude, interests, motivation level, and personality make up.

The study of all these variables again and again before the teaching of each unit will be wasteful, since many of them like intelligence, aptitude and personality make up can be assessed only at long intervals. Essentially teaching of a new unit should begin with what the student presently knows and should continue to where he is capable of the terminal performance under specified conditions and to a specified standard. According to DeCecco and Crawford (1977, p. 67):

> Different entering behaviour is required for different learning. Concept learning, for example, requires as entering behaviour the previous learning of verbal associations and discriminations, while rule learning requires as entering behaviour the previous learning of concepts.

After determining what entering behaviour is necessary for achieving a particular instructional objective, the teacher has to determine whether the student has acquired this behaviour. For this the teacher is required to construct a test which measures previous rather than present learning.

Instructional Procedures

After having entered the pupil behaviour, teacher has to select appropriate instructional procedures for providing instruction for various learning types and to ensure effective communication of subject matter. It includes the selection and development of appropriate software and hardware devices keeping in view the needs and level of students.

Evaluation

Finally, performance evaluation consists of tests and observation used to determine how well the student has achieved the instructional objectives. In case of any discrepancy, evaluation provides a feedback to other three compartments shown in the paradigm. Feedback from evaluation will show how well the student has achieved the instructional objectives, the adequacy of the entering behaviour, and the effectiveness of instructional procedures. The evaluation of achievement involves the use of absolute, relative, or criterion-related standards. Absolute and criterion-related standards describe the minimal level of mastery the student is expected to attain whereas the relative standard compares the performance of one student with others in his group. The former is concerned with the construction of criterion-referenced tests and latter with the construction of norm-referenced tests.

All the four compartments shown in the paradigm yield significant information and data for conducting research. Besides each one of them needs research so that more and more knowledge may be added to these areas and corrections may be made in case of any doubtful or ambiguous knowledge.

PRIORITY AREAS OF EDUCATIONAL RESEARCH IN INDIA

Educational research in India is of fairly recent origin and it is still in its infancy. It is because of the fact that the study of education as a subject at the university stage did not begin seriously until the forties. Although the need for educational research was acknowledged for the first time in the *Government Resolution on Educational Policy* *in 1913*, it is only after the attainment of Independence in 1947 a systematic effort to develop educational research was made in India. The research programmes in education received great stimulus in the first decade of free India by the establishment of national and state organizations for the development of elementary and secondary education, educational and vocational guidance, and research in curriculum and textbook development, and evaluation procedures. Certain organizations helped to spread research programmes by starting Extension Service Departments in a number of teachers' training colleges for conducting studies in a variety of techniques in service education of teachers. These organizations also financed certain research projects for improving curriculum, methods of teaching, and techniques of evaluation at the school stage.

Though the quantity of research work at the Master's and Doctoral levels has increased significantly over the years since Independence, its quality is poor or mediocre. The Education Commission (1966, p. 322) has attributed this to several reasons:

1. In the absence of specialized institutions, the bulk of research is being conducted by the students for Master's and Doctoral degrees. Most of this research is confined to training colleges which have very inadequate facilities for research and few competent people to guide it.
2. Very few scholarships are available for research students in education.
3. Ancillary services like documentation, computation, consultation, etc. have not been developed.
4. No central clearing house has been created and there has been considerable duplication of work.
5. The little research that has been done has largely remained in the archives, and administration has not used its findings for the formulation of policies.
6. The total expenditure on educational research has been negligible.

To develop educational research in quality and relate it effectively to the formulation of educational policies and improvement of education, the Education Commission (1966, pp. 322-323) has further recommended that:

1. A documentation centre and national clearing house in educational research should be developed at National Council of Educational Research and Training.
2. Educational research should be developed in teams and in interdisciplinary fields.
3. A National Academy of Education consisting of eminent educationists, broadly on the lines of the National Institute of Science, to promote educational thought and research should be set up.

4. The universities and research organizations should contribute in a big way to the development of educational research.
5. A good specialized training for research work and services for data processing, statistical analysis and consultation should be provided by universities and other research institutions.
6. It should be the responsibility of the NCERT at the national level and the State Institutes of Education at the state level to bridge the serious gap between educational research and current school practices. A similar role should be played by the UGC in the field of higher education.

Since the publication of the report of the Education Commission there has been an effective and systematic development of educational research. The NCERT and the UGC played a significant role in increasing awareness among the research workers of aims and techniques of educational research. Some journals devoted to educational research have been started by both the organizations. Material help in the form of financial assistance and equipment is also provided to individuals and institutions to encourage research at the school and university stages. Researchers or research groups, however, selected problems and topics they felt to be useful and conducted research on them without considering their place in the total pattern. Without such plans, educational research could not relate itself effectively to the formulation of educational policies and improvement of education. Thus when planning for educational research, we have to arrange the areas in order of priority.

The problem of determining priorities in educational research is complex and becomes essentially a subjective issue. Each researcher may have his priorities common or uncommon to others. These priorities may be assigned in terms of individual needs and capacities, social targets, finances, value patterns, consideration of time and funds and utility of a research problem. In this context, the Education Commission (1966, pp. 322-323) has noted that:

> The priorities in educational research at the national level, from the point of view of the educational planners or administrators, may be very different from similar priorities at lower levels, or from the point of view of teachers.

The Commission, therefore, did not favour evolving any rigid framework of priorities. It has suggested that the decisions about the priorities should be taken by the Education Research Council in the Ministry of Education at the national level, by the State Department of Education at the state level and by the universities, training colleges and teachers interested in the research work.

However, Third Survey of Research in Education (Buch, 1987) and Educational Research and Innovations Committee (ERIC) of the National Council of Educational Research and Training, New Delhi have suggested some priority areas in educational research keeping in view the National Policy of Education-1986: Programme of Action (1986) and the role of education in the total programme of national development.

1. Priority may be given to research studies designed to identify educational needs and problems of the scheduled caste and scheduled tribe children, rural girls, and children from economically deprived sections of the society. In this connection, studies on learning with interdisciplinary research, including bio-chemical and neuro-physiological factors, may be undertaken. Such studies will have immediate implications for the education of the children from the disadvantaged and deprived sections of the society.

2. Another priority area in the field of education is the fulfilment of the Constitutional Directive of Article 45 relating to the provision of free and compulsory elementary education for all children up to the age 14. Studies which will help to extend the educational programmes, formal and non-formal, to all children especially belonging to scheduled castes and scheduled tribes including nomadic, semi-nomadic and denotified communities; hill areas and slums; and economically weaker sections of the society may be conducted.

3. The drop out rate in school education is very high for various reasons and it is not likely to decrease very much in the near future. While every attempt should be made to increase the holding power of the school, we should also think of non-formal education as a possible alternative. Studies should be conducted to evolve organization and methods of self-learning so that children in school learn better and those who drop out from school are not left out of the country's educational effort.

4. The children from the scheduled castes, scheduled tribes, and the economically deprived sections of the society need educational opportunities as much as children from the other sections of society. How can opportunities be made equal? How can the effects of deprivation in the early life of infants be counteracted by social and educational intervention? These are some of the important problems which need the immediate attention of researchers.

5. The programme of vocationalisation of secondary education was initiated in 1976. Since then it has been implemented in 10 States and 5 Union Territories. A number of other states are also in the process of introducing vocationalisation. However, in spite of the

best efforts of central and state-governments, the scheme of vocationlisation of education has not yet picked up the desired momentum. There have been many factors responsible for the slow progress. Research relating to the study of such factors, namely, absence of a well co-ordinated management system, unemployability of vocational pass-outs, mis-match between demand and supply, reluctance in accepting the concept of vocationalisation by the society, absence of proper provisions for professional growth and career advancement for the vocational pass-outs etc. should be undertaken on the priority basis.

6. Keeping in view the rate of population growth, we will not be able to provide the necessary educational facilities to the increasing number of children. The number of schools and teacher will be so large that the necessary sources are not likely to be available. Research is needed to explore the new techniques and alternative strategies of education.

7. The system of education in the country should make significant contribution towards developing suitable educational opportunities for the physically, mentally and socially handicapped children. This needs special research efforts for designing suitable training and orientation programmes for teachers and developing suitable instructional aids and materials for teaching such type of children.

8. The identification and nurture of talent is another priority area of research in view of the rate of innovation and change in both technological and social spheres.

9. The scientific and technological developments have not been matched by corresponding advance in human character and values. Research is needed in value education for fostering universal and eternal values, and eliminating obscurantism, religious fanaticism, violence, superstition and fatalism.

10. Research is needed to explore how the existing educational structure and content can be altered, enhanced and modified to bring about faster economic growth and social change. In this connection research programmes on problems like life long/recurrent education, work and productivity oriented education, education and self-reliance, education and social mobility, strategies of human resource development, studies in educational finance, administration and planning including management of systems and cost effectiveness, need immediate attention of research workers.

11. The problems of the classroom, the teacher and the students, child development and processes of learning, and changing needs

should also find place among the priority areas of educational research.

POSTTEST

1. (*a*) Classify the fields of educational research in terms of the content areas of education.
   (b) List and discuss some significant research problems from these areas according to the needs of Indian education today.
2. (*a*) Name the different stages and sectors of education in India.
   (*b*) List and discuss some significant research problems relating to these stages and sectors of education.
3. List and discuss some significant research problems relating to teaching process: instructional objectives, entering pupil behaviour, instructional procedures and evaluation.
4. Describe some priority areas of educational research in India.

3. The Research Problem and Preparation of the Research Proposal

The six sections of this chapter reflect the following purposes. The first section which is quite brief, indicates that the researcher should be familiar with the fields or area of education from which he intends to choose his problem. The next three sections deal with the selection, definition, statement, and evaluation of the research problem. The fifth section describes the nature, importance and formulation of the hypotheses, the criteria of useable hypotheses, forms of stating, the hypotheses, and explains how the hypotheses are tested. The final section deals with the meaning and types of research proposal or synopsis, and the procedure followed in its preparation.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Describe why the researcher should be familiar with the field or area of education from which he intends to choose his problem.
2. List and describe the important sources for problem selection.
3. State what is meant by the 'definition' and 'statement' of the problem.
4. List and answer the questions which are helpful in the evaluation of a research problem.
5. State what is meant by hypothesis.
6. Discuss the importance of hypothesis.
7. List and describe the main criteria of usable hypotheses.
8. Give some examples of usable hypotheses.
9. List and describe the important conditions conducive to the formulation of hypotheses.
10. List and describe various forms of hypothesis.
11. Give some examples of the various forms of hypothesis.
12. State how the hypotheses are tested.
13. Explain the meaning of a research proposal or synopsis.
14. Describe various types of research proposal or synopsis.
15. Describe the procedure used in the preparation of a research proposal or synopsis.

The first step in the research process is the choice of a suitable problem for investigation. Research starts with a felt difficulty. It takes place when there is a problematic situation and a need to solve the problem.

In each field or area of educational research several problems exist which may have reference to the pure, applied, or action research. It is essential to identify such problems. solve them and thereby enrich the corpus of knowledge constituting the science of education.

The identification of a research problem is a difficult but an important phase of the entire research process. It requires a great deal of time, energy, and logical thinking on the part of the researcher. A beginning researcher finds the task of identifying a research problem a difficult one. This may be due to his limited knowledge of research process and his unpreparedness for identifying the problem. He may also be unfamiliar with the areas in which research is needed and the procedures he is to follow for selecting a suitable field for research.

KNOWING THE FIELD

It is first necessary for a researcher to choose a broad field within which he will conduct the study. The researcher should have a deep interest in the field. His enthusiasm and real curiosity enhance his motivation for the successful completion of the study. The field selected should be such in which the researcher is capable of demonstrating necessary initiative, originality, and good judgment.

A thorough understanding of the known facts and ideas in the field or area in which the researchier is interested constitutes the first and most important step in selecting a problem for study. If the researcher is familiar with his field and knows what studies have already been conducted in it, he will also know about the problems which have remained unsolved. A survey of suggestions for further research given at the end of research reports and reviews of research is helpful to the researcher in getting an idea about the gaps which exist is knowledge in a particular field of educational research. Periodicals and annual bibliographies which appear in most fields of educational research are helpful in keeping the researcher informed about the research going on in the field in which he has competence and interest.

SELECTION OF THE PROBLEM

After selecting the broad area, the researcher must narrow it down to a highly specific research problem. He must state the specific question whose answer he seeks through the application of scientific method.

It is not possible to list all the educational problems that are able to be and need to be researched. Each researcher selects a problem because of his own unique needs and purposes. There are, however, some important sources which are helpful to a researcher for selecting a problem.

Professional Experience

One of the most fruitful sources of problems for the beginning researcher are his own experiences as a professional educator. In the classroom there is a dynami interaction between teacher and pupil, between pupil and pupil, and between pupil and materials. This interaction provides a rich source of problems to be solved through educational research. Classrooms lectures and discussions suggest many stimulating problems to be solved. The teacher may be confronted with a number of behavioural problems in the classroom. He has to make decisions about the probable effects of classroom instruction on pupil behaviour so as to establish a relationship between instructional objectives, learning experiences and pupil change. The pupil behaviour may be analysed in terms of his achievement, interest, attitudes, intelligence, aptitudes, motivation, and personality. The teacher has to make decisions about teaching methods and techniques. He may study the effectiveness of a particular method of teaching so as to base classroom instruction upon empirical evidence rather than upon his biased impressions. For instance, a teacher in a higher secondary school may wish to investigate whether the problem-discussion method or the lecture method is more effective in the teaching of civics to eleventh class students.

Similarly, the teacher has to decide about the instructional objectives for teaching each unit of work and then to evaluate whether the instructional objectives has been realized. In case of any discrepancy, the teacher may wish to investigate the drawbacks in the formulation of objectives, planning or/and organization of learning experiences, and the process of evaluation itself.

The classroom is not the only source of problems for the teacher. The experiences of the pupils, their attitudes, home environment, peer influences, socio-economic status, and motivation level provide rich sources of rich problems.

The educational administrators may find subjects of research in the areas concerning decision-making, scheduling, teacher placement, instructional supervision, and several other-matters with which they are concerned. Different professional personnel in the field of education view the educational scene from different angles and they may encounter many problems of research in their own perspective.

Contacts and discussions with research oriented people, attending conferences, seminars and listening to the learned speakers are helpful in locating research problems. Active membership in organizations which are concerned with the improvement of educational system usually brings one into close contact with crucial problems and issues concerning education.

Inference From Theory

A second important source of research problem lies in the inferences that can be drawn from various educational and psychological theories known to the researcher. The application of general principles involved in various theories to specific classroom situation makes an excellent starting point for research. It will help to determine whether a particular theory can be translated into actual practice. Learning theories, personality theories, theories of intelligence, theories of motivation, sociological theories and many others make an excellent starting point for research in classroom situations. It is only through the application of scientific method, we can profitably test the validity, scope, and practicability of various theories in educational situations.

Professional Literature

Among the sources of problems one has to be thoroughly acquainted with the literature in the field of one's interest. The study of professional literature will not only expose a researcher to pressing research problems but will suggest the way in which research is conducted. Research reports, bibliographies of books and article, periodicals, research abstracts, year books, dictionaries and research guides suggest areas that need research. Some specialized sources as the *Encyclopaedia of Educational Research*, *Dissertation Abstracts International*, the *Hand-book of Research on Teaching*, *Psychological Abstracts* and similar publications provide rich sources of problems. The publications like *Research Needs in the Study of Education* (1968) and the *Third Indian Year Book of Education : Educational Research* (1968) are exclusively devoted to identifying and bringing into sharp focus the varied research needs in the different areas and aspects of Indian education. All published research reports generally conclude by making suggestions about further research. Such suggestions help a researcher to consider how the procedures employed could be adapted to solving other problems, or how a similar study could be undertaken in a different field or subject area or with a different sample or subjects.

Replication or extension of completed research studies is also a profitable and worthwhile activity for a beginning researcher. There is a common feeling that once a study has been conducted it should not be repeated. Repeating a study at different times on different groups and in different places increases the extent to which the research findings can be generalized and provides additional evidence of the validity of the findings.

Sometimes research studies are criticized for weakness in design, tools, treatment and analysis of data, contradictions and inconsistencies in the results, and so on. A valuable and worthwhile contribution can be made by repeating these studies by making necessary modifications in the designs and procedures so as to correct the findings for their inconsistencies.

Technological and Social Changes

Technological and social changes demand development of new courses and curriculum for students in educational institutions. All these developments constantly bring forth new problems for research. The use of hardware and software in classroom instruction, the training of teachers in the methodology of teaching through team teaching, micro-teaching, simulation, etc. are being advocated by the educationists all over the country. All such innovations in education need to be carefully evaluated through research process.

DEFINITION AND STATEMENT OF THE PROBLEM

After the problem has been selected the next task is to define it in a form amenable to research. According to Whiteny (1964, pp. 80-81):

> To define a problem means to put a fence round it, to separate it by careful distinctions from like questions found in related situations of need.

The researcher must be certain that he knows exactly what his problem is before he begins work on it. Specifying a problem explicitily and narrowing it down to workable size are extremely important at the very start. According to Monroe and Engelhart (1928):

> To define a problem means to specify it in detail and with precision. Each question and subordinate question to be answered is to be specified. The

> limits of the investigation must be determined. Frequently, it is necessary to review previous studies in order to determine just what is to be done. Sometimes it is necessary to formulate the point of view or educational theory on which the investigation is to be based. If certain assumptions are maue, they must be explicity noted.

This quotation describes the means of easing the burdens of research by a clear and precise definition of the problem.

The following rules, according to Hillway (1964, p. 117) will be found useful by a researcher in setting down on paper the final definition of a problem:

1. Be sure that the topic chosen is neither too vague nor too broad in scope.
2. To make the problem clearer and more understandable, state it as question which requires a definite answer.
3. Carefully state the limits of the problem, eliminating all aspects and factors which will not be considered in the study.
4. Define any special terms that must be used in the statement of the problem.

In the formal definition of the problem the researcher is required to describe the background of the study, its theoretical basis and underlying assumptions, and state the problem in concrete, specific, and workable questions.

All questions raised must be related to the problem. Each major issue or element should be separated into its subsidiary or secondary elements, and these should be arranged in logical order under the major division.

A good statement of a problem must clarify exactly what is to be determined or solved. It must restrict the scope of the study to specific and workable research questions. The most important step in this direction is to specify the variables involved in the questions and define them in operational terms.

In stating a problem, the researcher should see that it is neither stated in so general terms as to become vague, with no clear indications of the direction the study is to take, nor specified so narrowly as to become insignificant and trivial.

To illustrate, a researcher states that he is interested in studying the effectiveness of programmed instruction on the performance in Algebra of high school students. This statement is broad and it communicates in a general way what the researcher wants do to. However, it is needed to specify the problem with much greater clarity if the researcher is to find a method to arrive at the right type of conclusions.

The first step is to specify the variables involved in the problem and define them in operational terms. The variable involved in the problem are "effectiveness", "programmed instruction", "performance", and "high school students". It may be noted that these variables are not given dictionary meanings. For example, the dictionary meaning of "effectiveness" is "producing the desired effect". This meaning is not sufficient for research purposes. It is important for a researcher to specify exactly what indicator of effectiveness he will use or what he will do to measure the presence or absence of the phenomenon denoted by the term "effectiveness". Similarly the researcher has to define other variables also in terms of the operations or processes that will be used to measure them. To define operationally, it is always essential to designate some kind of over behaviour that is directly observable and measurable by the researcher and others to represent those concepts involved in the variables.

In this study, the researcher might choose to define "effectiveness" as the improvement made by the students in scores on a standarized criterion test of Algebra. The term "programmed instruction" would be defined as the programmed text in Algebra. The term "performance" may be defined as scores on a criterion test in Algebra. High school might refer to those schools that run classes upto tenth and are affiliated to a particular Board of School Education. The operational definitions of the variables help to narrow the scope of a general question to specific measurable and observable variables.

There is no generel agreement whether a problem be presented as a question or in the form of a statement. However, the question form is preferred since it is simple and straightforward, and it guides the researcher to the task at hand.

The problem should be statad in such a way that it indicates a relationship between two or more variables. It should neither involve philosophical issues nor values or judgmental questions that cannot be answered by scientific investigation.

EVALUATION OF THE PROBLEM

When considering a problem which a researcher may undertake for investigation, he is required to ask himself a series of questions about it. These questions are helpful in the evaluation of the problem on the basis of personal suitability of the researcher and that on social value of the problem. All these questions must be answered affirmatively before the study is undertaken.

1. Is the Problem Researchable?

There are certain problems that cannot be effectively solved through the process of research. The question such as, "Is it good to provide sex education in the secondary schools?" is a value question and can be answered only on the basis of value judgment. The educationists, parents, students, and teachers may provide answer to such type of questions, but in practice questions involving philosophical issues cannot be answered by scientific investigations. Science cannot provide answer to philosophical and ethical questions. A reasonable problem is always concerned with the relationship existing between two or more variables that can be defined and measured.

From the foregoing discussion it may not be assumed that research cannot be used in developing solutions to philosophical and ethical issues. The point that is being emphasized here is that the problem must be stated in workable research questions that can be answered empirically. For instance, the above question might be restated as, "what is the effect of sex education in the secondary schools on the attitudes of adolescent girls towards premarital indulgence in sex?" The researcher can design a study to obtain information on this type of question that can be used in developing a solution to the ethical question, "Is it good to provide sex education in the secondary schools?"

2. Is the Problem New?

There is no purpose in studying a problem which had already been adequately investigated by other researchers. In ignorance a researcher undertakes a study which is neither new nor original. He, therefore, spends time needlessly on a problem already investigated by some other researcher. To avoid such duplication, it is essential to examine very carefully the record of previous studies completed in one's field. The researcher should not select a problem until he is convinced that it really is a new problem which has never before been investigated successfully. However, this does not mean that a problem which has been investigated in the past is no longer worthy of study. A researcher may repeat a study when he wants to verify its conclusions or to extend the validity of its findings in a situation entirely different from the previous one.

3. Is the Problem Significant?

The question of the significance of the problem usually relates to what a researcher hopes to accomplish in a particular study. What is his purpose in undertaking to solve the particular problem he has chosen? What new knowledge does he hope to add to the sum total of what is known? And what value is this new knowledge likely to have? All these questions are directed to the researcher. Unless these questions can be answered clearly by him, the problem should not be selected for research.

The researcher should show that the study is likely to fill in the gaps in the existing knowledge, to help resolve some of the inconsistencies in previous research, or to help in the reinterpretation of the known facts. The findings should become a basis for theory, generalizations or principles and should lead to new problems for further research. However, if the findings lack apparent implications, then they must, at least, have some practical application.

4. Is the Problem Feasible for the Particular Researcher?

A problem may be a good one from the point of view of the three criteria mentioned earlier, but yet it may not be feasible in view of some of the personal aspects of a researcher discsssed below:

(*a*) *Research competencies.* The problem should be in an area in which the researcher is qualified and competent. He must be familiar with the existing theories, concepts and laws in order to identify a worthwhile problem. He must also possess the necessary skills and competencies that may be needed to develop, administer, and interpret the necessary data-gathering tools. Furthermore, he needs to consider whether he has the necessary knowledge of research design and statistical procedure that may be required to carry out the research through its completion.

(*b*) *Interest and enthusiasm.* The problem should be one in which the researcher is genuinely interested and about which he is truly enthusiastic. It should be meaningful and arouse real curiosity of the researcher. The investigation of a problem in which the researcher has little or no interest can prove the most tiresome kind of distasteful task.

(*c*) *Financial considerations.* The problem should be one which is financially feasible. The researcher should ascertain whether he has the necessary financial resources to carry on the investigation of the selected problem. He must have an estimate of the expenditure involved in data-gathering equipment, printing, test materials, travel, and clerical assistance. If the research project is an expensive one, the researcher may determine the possibility of getting financial assistance from the state and private agencies and other research organizations like the University Grants Commission, the National Council of

Educational Research and Training and the Indian Council of Social Sciences Research.

(*d*) *Time requirements*. The problem should be one that can be studied and completed in the allotted time. The researcher should select a problem that can be investigated with little expenditure of time or energy. It is worthwhile to plan for the time that will be needed for the development and administration of tools, processing and analysis of data, and writing of the research report. If the problem is a big one, then it is essential to explore the possibilities of seeking cooperation from others and taking it up as a team project.

(*e*) *Administrative considerations*. In addition to personal, financial and time requirements, the researcher should consider the kinds of data, equipment, specialized personnel, and administrative facilities that are needed to complete the study successfully. The researcher must ascertain whether the pertinent data are available and accessible to him. He must also check whether the school authorities permit him to contact the students, administer the necessary tools or conduct necessary experiments, interview the teachers, or have access to important records and documents.

THE HYPOTHESES

Once the selection and definition of the problem have been accomplished, the derivation of hypotheses is the most important step in the research process.

Etymologically *hypothesis* is made up of two words, "hypo" (less than) and "thesis", which mean less than or less certain than a thesis. It is the presumptive statement of a proposition or a reasonable guess, based upon the available evidence, which the researcher seeks to prove through his study. The terms hypothesis, theory, and conclusion occur frequently in research literature and need some clarification. The *hypothesis* is precisely defined as a tentative or working proposition suggested as a solution to a problem, and the *theory* as the final hypothesis which is defensibly supported by all the evidence. The final hypothesis which fits all the evidence becomes the chief *conclusion* inferred from the study. According to Hilway (1964, p. 123):

> From the research scholar's point of view, the hypothesis may be conveniently considered as a tentative or working assumption, and the theory as the surviving or final hypothesis which is most defensibly supported by all the evidence. But since knowledge arrived at through the scientific method is subject to revision in the light of new data, a theory is in only one sense always only a working assumption, so that the conventional distinction between hypothesis

> and theory (on the basis of increasing adequacy of evidence and hence of greater certainty) is a relative one. For all practical purposes, the duality of the term hypothesis need not concern the research scholar except to serve as a salutary reminder that no conclusion reached by the scientific method is necessarily a fixed and final truth.

The hypothesis is a powerful tool in research process to achieve dependable knowledge. It helps the researcher to relate theory to observation and observation to theory.

Hypotheses are formulated only as the suggested solution to the problem, with the objective that the ensuing study may lead either to its rejection or to its retention. They enable the researcher to locate and identify the variables involved in the study and suggest methodological procedures that are to be employed.

Importance of the Hypotheses

The importance of hypotheses is generally recognized more in the studies which aim to make predictions about some outcome. In experimental research, the researcher is interested in making predictions about the outcome of the experiment or what the results are expected to show and therefore the role of hypotheses is considered to be of utmost importance. In the historical or descriptive research, on the other hand, the researcher is investigating the history of a city or nation, the life of a man, the happening of an event, or is seeking facts to determine the status quo of some situation and thus may not have a basis for making a prediction of results. A hypothesis, therefore, may not be required in such fact-finding studies. Hillway (1964, p. 130) too is of the view that "when fact-finding alone is the aim of the study, a hypothesis may not be required."

Most historical or descriptive studies, however, involve not only fact-finding but interpretation of facts to draw generalizations. If a researcher is tracing the history of a school or making a study about the results of a coming assembly poll, the facts or data he gathers will prove useful only if he is able to draw generalizations from them. Whenever possible, a hypothesis is recommended for all major studies to explain observed facts, conditions, or behaviour and to serve as a guide in the research process. If the hypotheses are not constructed, a researcher may waste much time and energy in gathering extensive empiricial data and then find that he cannot state facts clearly and detect relevant relationships between variables as there is no hypothesis to guide him. The importance of hypotheses may be summarized as under (Ary *et al.* 1972, pp. 73-74) :

1. Hypotheses facilitate the extension of knowledge in an area. They provide tentative explanations of facts and phenomena, and can be tested and validated. Such explanations, if held valid, lead to generalizations which help significantly in understanding a problem and thereby extend the existing knowledge in the area to which they pertain.
2. Hypotheses provide the researcher with rational statements, consisting of elements expressed in a logical order of relationships which seek to describe or to explain conditions or events, that have not yet been confirmed by facts. Some relationships between the elements or variables in hypotheses are known facts and others transcend the known facts to give reasonable explanations for known conditions. The hypotheses enable the researcher to relate logically known facts to intelligent guesses about unknown conditions.
3. Hypotheses provide direction to the research. They represent specific objectives and thus help the researcher to determine the type of data needed to test the proposition. The hypotheses tell the researcher specifically what he needs to do and find out in his study. They help in the selection of relevant facts and variables that the researcher needs in his study. Hypotheses provide a basis for selecting the sample and the research procedures to be used in the study. The statistical techniques needed in the analysis of data, and the relationships between the variables to be tested, are also implied by the hypotheses. Furthermore, the hypotheses help the researcher to dclimit his study in scope so that it does not become broad or unwieldy.

For example, a researcher is interested in studying the effect of test anxiety on the academic achievement of sixth grade children. He hypothesizes: "There is a negative relationship between the test anxiety and academic achievement of sixth grade student." It is clear from the statement of the hypothesis that the researcher would use an experimental method in which he may compare the academic achievement of a group of sixth grade students with high text anxiety with the achievement of a similar sample of students with low test anxiety. The significance of the difference between the mean academic achievement scores of the two groups could be analyzed by using t-test or analysis of variance technique.

4. Hypotheses provide the bases for reporting the conclusions of the study. The researcher will find it very convenient to test each hypothesis separately and state the conclusions that are relevant to each. On the basis of these conclusions, he can make the research report interesting and meaningful to the reader.

Formulation of the Hypotheses

Hypotheses are guesses or tentative generalizations, but these guesses are not merely accidents. Collection of factual information alone does not lead to successful formulation of hypotheses. Hypotheses are the products of considerable speculation and imaginative guess work. They are based partly on known facts and explanations, and partly conceptual. There are no precise rules for formulating hypotheses and deducing consequences from them that can be empirically verified. However, there are certain necessary conditions that are conducive to their formulation.

1. *Richness of background knowledge.* A researcher may deduce hypotheses inductively after making observations of behaviour, noticing trends or probable relationships. For example, a classroom teacher daily observes student behaviour. On the basis of his experience and his knowledge of behaviour in a school situation, the teacher may attempt to relate the behaviour of students to his own, to his teaching methods, to changes in the school environment, and so on. From these observed relationships, the teacher may inductively formulate a hypothesis that attempts to explain such relationships.

Background knowledge, however, is essential for perceiving relationships among the variables and to determine what findings other researchers have reported on the problem under study. One finds limitless knowledge in almost any area of human pursuit which has accumulated in the form of literature as a result of the painstaking efforts of generation. New knowledge, new discoveries, and new inventions should always form a continuity with the already existing corpus of knowledge and, therefore, it becomes all the more essential to be well versed with the already existing knowledge. The researcher may be led astray, if he does not have thorough acquaintance with and mastery of the existing knowledge.

Relevant to a particular problem, various researchers may have different hypotheses but the soundness of the hypotheses will depend upon the quantum of knowledge that one possesses of the area of investigation. Significant researches have not taken place accidently and if it be a mere accident to come across some important discovery, the researcher must have the rich background knowledge which may enable him to perceive that important discovery. Closed minds tend to see problems from a narrow and biased viewpoint but liberal and objective educated minds look at them from multiple points of view suggested by their multiple experiences. Hypotheses may be formulated correctly by persons who have enriched experiences and academic background but they can never be formulated by those who have poor background knowledge.

2. *Versatility of intellect.* Hypotheses are also derived through deductive reasoning from a theory. Such hypotheses are called deduc-

tive hypotheses. A researcher may begin a study by selecting one of the theories in his own area of interest. After selecting the particular theory, the researcher proceeds to deduce a hypothesis from this theory through symbolic logic or mathematics. This is possible only when the researcher has a versatile intellect and can make use of it for restructuring his experiences. Creative imagination is the product of an adventure, sound attitude and agile intellect. In the hypotheses formulation, the researcher works on numerous paths and he may be led to blind alleys and only rarely he may be led to the right path. The researcher, therefore, has to exercise a great restraint and command considerable patience. He has to make a consistent effort and develop certain habits and attitudes. Moreover, the researcher has to saturate himself with all possible information about the problem and then think liberally at it and proceed further in the conduct of the study.

3. *Analogy and other practices.* Analogies also lead the researcher to clues that he might find useful in the formulation of hypotheses and for finding solutions to problems. For example, suppose a new situation resembles an old situation in regard to a factor X. If the researcher knows from previous experience that the old situation is related to other factors Y and Z as well as to X, he reasons that perhaps a new situation is also related to Y and Z. The researcher, however, should use analogies with caution as they are not fool proof tools for finding solutions to problems.

At moments conversations and consultations with colleagues and experts from different fields are helpful in formulating significant hypotheses.

Criteria of Usable Hypotheses

It is essential that hypotheses be carefully formulated. Travers (1978, pp. 76-78) has suggested seven criteria to which they should conform:

1. *Hypotheses should be clearly and precisely stated.* When hypotheses are clearly stated they usually avoid the use of general terms such as personality, intelligence, social class, etc. in their statements. On the other hand, the researcher may use "personality as measured by the Sixteen Personality Factor Questionnaire," "Intelligence as measured by Raven's Progressive Matries," or "social class as defined by socio-economic status scale by Jalota *et al*." The clear statement of hypotheses generally involves concise technical language and definition of terms that are better defined than those in common language.

2. *Hypotheses should be testable.* Hypotheses should be formulated

in such a way that they can be tested or verified. Such hypotheses enable the researcher to determine by observation whether those consequences that are derived deductively, actually do or do not occur. If the hypotheses are not testable, it would be impossible either to confirm or contradict them and therefore they do not help the researcher to draw conclusions. For example, the hypothesis: "The N.C.C. programme promotes the all-round adjustment of the high school students" would be hard to test because of the difficulty of defining and measuring "all-round adjustment." Moreover, it would be difficult for the researcher to isolate other factors that might contribute to the adjustment of high school students.

Since hypotheses are predictors of the outcome of the study, they must relate variables that are capable of being measured. It is essential that the tools or instruments should exist (or can be developed) which will provide measures of the variables involved. If no tools or means are available for measuring the variables, then it would be impossible for the researcher to collect the data necessary to test the validity of the hypotheses. A hypothesis, therefore should define the variables operationally by stating the operations or procedures necessary to measure them. For example, the hypothesis: "There is a negative relationship between neuroticism and achievement in mathematics of fifth grade students" meet the criterion of testability because the variables involved in the statement can be defined operationally. Neuroticism might be defined as "the scores made on the neuroticism-stability scale of Maudsly Personality Inventory" and achievement in mathematics defined as "marks obtained in mathematics by the fifth grade students in their annual examination."

3. *Hypotheses should state the expected relationship between variables.* A satisfactory hypothesis should state explicitly an expected relationship between the variables. Let us consider two hypotheses:

(*i*) High school students who attend N.S.S. programmes show greater moral growth than students who do not.

(*ii*) Extraversion, as measured by the Maudsly Personality Inventory, will be related positively to achievement in social studies of seventh grade students, as measured by standardized test X.

The first hypothesis is not usable because the term "greater moral growth" does not refer to a variable that is measurable at the present time, or likely to be measured in the near future. On the other hand, the second hypothesis refers to the variables 'extraversion' and 'attainment in social studies' that can be measured by Maudsly Personality Inventory and the standardized test X respectively. Moreover, the relationship posited is that the personality variable and the attainment variable will be related positively.

4. *Hypotheses should be limited in scope* Hypotheses of global significance are not usable as they are not specific and simple for testing and drawing conclusions. A beginning researcher, however, is overly ambitious in his initial efforts and formulates hypotheses of global significance. It is partly because of his earnestness and partly because it takes maturity of viewpoint to realize how little can be accomplished in a specified period. It is desirable to formulate hypotheses that are simple to test, and yet are highly significant Sometimes it is of course possible to state a rather broad research hypothesis and derive a number of operational hypotheses from it.

5. *Hypotheses should be consistent with most known facts.* Hypotheses should not be inconsistent with a substantial body of established fact. They should be grounded in the well established theories and laws. Consider the hypotheses: "There is no relationship between the self concept of adolescent male students and their rate of physical growth." This hypotheses is not worth testing because the preponderance of evidence supports the relationship between self concept and rate of physical growth.

Hypotheses, however, cannot be consistent with all known facts because in so many areas the facts themselves contradict one another. In such cases it is worthwhile to formulate hypotheses that resolve the contradiction.

6. *Hypotheses should be stated as for as possible in simple terms.* Stating the hypotheses in simple terms not only makes their meaning clear to others, but also helps in their testability. Moreover, the simplicity of statement provides a basis for a clear and easily comprehended report at the completion of the study.

Hypotheses should not make use of the vague terms or constructs. It is quite useless to formulate a hypothesis that makes use of the terms or constructs which do not convey the intended meaning to the reader. The researcher should make use of such terms that are generally accepted for naming a phenomenon.

7. *The hypotheses selected should be amenable to testing within a reasonable time.* The researcher should not select a problem which involves hypotheses that are not amenable to testing within reasonable specified time. He must know that there are problems that cannot be solved for a long time to come. These are problems of immense difficulty that cannot be profitably studied because of the lack of essential techniques or measures.

Statement of the Hypotheses

A research or substantive hypothesis must be stated in a testable form for its proper evaluation. As already stressed this form should indicate a relationship between the variables in clear, concise, and understandable language.

Research hypotheses are classified as being *directional* or *non-directional*. The hypotheses which stipulate the direction of the expected differences or relationships are termed as directional hypotheses. For example, the research hypotheses: "There will be a positive relationship between individual's attitude towards high caste Hindus and his socio-economic status," is a directional research hypotheses. This hypothesis stipulates that individuals with favourable attitude towards high caste Hindus will generally come from higher socio-economic Hindu families and therefore it does stipulate the direction of the relationship. Similarly, the hypotheses: "Adolescent boys with high IQ will exhibit low anxiety than adolescent boys with low IQ" is a directional research hypothesis because it stipulates the direction of the difference between groups.

A research hypothesis which does not specify the direction of expected differences or relationships is a non-directional research hypothesis. For example, the hypothesis: "There will be difference in the adaptability of fathers and mothers towards rearing of their children" or "There is a difference in the anxiety level of adolescent girls of high IQ and low IQ" are non-directional research hypotheses. Although these hypotheses stipulate there will be a difference, the direction of the difference is not specified.

A research hypothesis can take either *declarative* form, the *null* form, or the *question* form.

When the researcher makes a positive statement about the outcome of the study, the hypotheses takes the declarative form. For example, the hypothesis: "The academic achievement of extroverts is significantly higher than that of the introverts," is stated in the declarative form. In such a statement of hypothesis, the researcher makes a prediction based on his theoretical formulations of what should happen if the explanations of the behaviour he has given in his theory are correct.

In the null form, the researcher makes a statement that no relationship exists. The hypothesis, "There is no significant difference between the academic achievement of high school athletes and that of non-athletes," is an example of null hypothesis. Since null hypothesis can be tested statistically, they are often termed as *statistical* hypotheses.

consequences.

The necessary conditions for confirmation are :

(a) All factual evidence collected through tests or other means should correspond with the deduced consequences.

(b) The test situation or data collecting tool should take into account all factors and conditions that are suggested by the consequences.

(c) The consequences are logically deduced from the hypotheses.

2. The absence of conflict with the other satisfactory proved generalizations lends support to the correctness of a hypothesis.

3. A hypothesis is also confirmed to be correct if the predictions made on its basis prove to be successful.

It is important for the researcher to formulate hypotheses before data are gathered. This is necessary for an objective and unbiased study.

THE RESEARCH PROPOSAL OF SYNOPSIS

(After the formulation of hypotheses, the researcher should write out a proposal, synopsis, or plan for research including a statement of the problem and hypotheses, the research procedures, the sample to be selected, the tools to be selected or developed for the collection of data, and the statistical techniques to be used for the analysis and interpretation of data.)

The preparation of a research proposal is a prerequisite in the research process. In many institutions a researcher is required to submit a proposal of the research project for its approval. This serves as a basis for determining the feasibility of the project and provides a systematic plan of procedure for the researcher to follow. It also gives the research supervisor (if any) a basis for guiding the researcher in the conduct of research study.

The research proposal is a systematic plan, which brings to focus the preliminary planning that will be needed to accomplish the purpose of the proposed study. It is just like a blueprint which the architect prepares before the construction of a building starts. The initial draft of the proposal is subject to modification in the light of the constructive criticism, comments and suggestions of the experts in the field. These experts include researcher's supervisor, colleagues, research and statistical specialists, or other experienced researchers who act as members of the research committee of an institution or organization. The research proposal always goes through the developmental phases and therefore, deficiencies, if any may be corrected by discussing the same with the experts. An acceptable research proposal, therefore, is usually the result of several modifications and as such there should not be any haste on the part of a researcher in its preparation.

Types of Research Proposal

Research proposal can be broadly classified into three groups. This classification is based mainly on the type of the research study to be conducted and the purpose for conducting it.

1. The most common proposal is the type that is prepared by those who wish to conduct research for their master's degree in education or their doctoral degree in education. Submission of a research proposal is a general requirement of most of the universities for beginning researchers. The proposal is evaluated by a committee of experts to determine the feasibility of the study.

2. The second type of research proposal is one that is submitted by a researcher to a private or governmental agency for the financial assistance. Such an assistance is given on the recommendation of panel of experts after evaluating the proposal.

3. Finally, in the third category is the research proposal that is submitted by an university or a college teacher to some research organization, such as university, Indian Council of Social Sciences Research, National Council of Educational Research and Training, University Grants Commission, or any other autonomous research organisation. Since more funds at present are earmarked by such organizations for research, more teachers are taking interest to investigate problems in the areas of concern in the field of education at the school, college or university levels.

Preparation of the Research Proposal

There are certain specifications that are essential to good research and, therefore, all such specifications should appear in each plan. A worthwhile research study is likely to result only from a carefully planned and well-designed proposal. The following categories of information should appear in the research proposal.

INTRODUCTION

The introductory part of the proposal should include the following information:

1. The Title

The title of the research proposal should do no more than name the topic. It should be so worded that it suggests the theme of the study. In selecting a title, the researcher should consider two things:

(*a*) The title should not be burdened by pompous words and should not include terms of unscientific; rhetorical, argumentative, emotional, or biased nature. The language in the title should be professional in nature but not pedantic.

(*b*) The title should not be too lengthy or too involved. It should be specific to the area of study. An adequate title for a study is one which represents a reasonable restriction and implies an objective approach. It must be noted that the proposed study cannot completely be presented in the title; but it should be so worded that it would give sufficient information about the nature of the study.

2. Statement of the Problem

Statement of the problem is not exactly the same as the title of the thesis. It has a definite place in the introductory chapter and is an attempt to focus on a clear goal. Statement of the problem should primarily be an expansion of the title. It should be either in question form or as a declarative statement. The major question or statement may be followed by several minor questions or statements.

3. Review of Related Literature

The theoretical and empirical framework from which the problem arises must be briefly described. Both conceptual and research literature are to be reviewed for this purpose. The latest research trends pertinent to the problem should also be mentioned in this section. The researcher has to make it clear that his problem has roots in the existing literature but it needs further research and exploration. A brief resume of related studies found in journals, magazines, abstracts, and reports should be made. This provides evidence that the researcher is familiar with what is already known and also with what is unknown and unproved. An analysis of the previous research eliminates the risk of duplication of what has been done, and provides bases for formulating hypotheses.

4. The Hypotheses

Questions that the research is designed to answer are usually framed as hypotheses to be tested on the basis of evidence. This step establishes the problem and the logic underlying the research study. It gives direction to the data gathering procedure. It is in the light of the hypotheses that the relevance of data to be collected is judged.

The formulation of the research hypotheses typically is determined by the implications of the related literature and the deductive logic of the problem under investigation. They should meet all the seven criteria discussed earlier.

5. Significance of the Study

A research proposal should justify the worthwhileness and urgency of the research study. It should indicate clearly how the results of the research can influence educational theory or practice. The need for undertaking a research study can be shown in several ways.

One way of establishing a need for a research study is to show the time lapse between the earlier study and the present one; and therefore the new knowledge, techniques, or conditions indicate a need to replicate the study.

A second way of showing the need for a study is to divulge that there are gaps in the knowledge provided by previous research studies and to show how the present study will help to fill in these gaps and add to the quantum of existing knowledge. ✓

A third and final way to accomplish the need for a study is to show the lack of information about a problem by presenting the supporting statements of other research studies.

6. Definition of Terms and Coneepts

It is necessary to define all unusual terms and concepts that could be misinterpreted. The technical terms or words and phrases having special meanings need to be defined operationally. We seldom come across unanimous definition or modes of measurement for certain concepts and terms, such as intelligence, motivation, achievement, etc., and in such situations it becomes obligatory on the part of the researcher to provide an operational definition of such terms or concepts by stating how the variables will be measured and interpreted. The *Dictionary of Education*, edited by Carter V. Good; the *Comprehensive Dictionary of Psychological and Psychoanalytical Terms*, edited by Horace B. English and Eva C. English; and the *Dictionary of Sociology*, edited by H.P. Fairchild are valuable sources for definitions of technical terms.

7. Delimitations of the Study

Boundaries of the study should be made clear with reference to (*a*) the scope of the study by specifying the areas to which the conclusions will be confined, and (*b*) the procedural treatment including the sampling procedures, the techniques of data collection and analysis, the development of measuring tools and their use in the study.

8. Basic Assumptions

Assumptions are statements of ideas that are accepted as true. They serve as the foundation upon which the research study is based. According to Turney and Robb (1971, p. 29) :

> They are written to reveal what the researcher, and others, theorize to be true in the study to be conducted; moreover, they are also written to indicate those factors affecting the study that are necessarily unalterable.

The researchers in writing their proposals should select their basic assumptions with care and be particularly aware of certain common errors. Turney and Robb (1971, p. 30) have listed these errors as under:

1. The first error is failure to state all of the basic assumptions that are relevant to the study. This situation usually occurs in questionnaire studies, in which necessary assumptions are taken for granted without being designated as basic assumptions of the study.

2. The second error concerns irrelevant assumptions. Such assumptions are neither applicable nor necessary to the study.

3. The third error is that of selecting unsubstantiated assumptions which cannot be defended by any logic, empirical evidence (objective data), or authoritative sources.

PROCEDURES FOR COLLECTING DATA

In this section the details about sampling procedure and the data collecting tools are described. It may be noted that in historical and philosophical researches the nature of the data and their treatment is different from other types of researches.

1. Sampling

In research situations the researcher usually comes across unmanageable populations, wherein large numbers are involved. In such cases the researcher may use different sampling methods, such as random, systematic, multistage, stratified, cluster, judgement, or quota sampling depending upon his requirement. The researcher should be well informed about the scope and limitations of each one of these sampling methods before he thinks of his own sample in the research. A research proposal should clearly define the population from which the researcher will draw his sample, and describe the procedure he will use to select the sample.

2. Tools

In order to collect evidence or data for a study, the researcher has to make use of certain testing and non-testing tools such as intelligence tests, aptitude tests, achievement tests, personality inventories, observation, interview, questionnaire, rating scales, etc. The researcher must be well versed in the use of these tools and fully aware of their merits and demerits.

The research proposal should explain the reasons for selecting a particular tool or tools for collecting data. The proposal should also include the reported evidence of the reliability and validity of the standardized tools for the purpose of the study. The researcher should take help of *Mental Measurement Year Books* (*Buros*), *Indian Mental Measurement Year Book* (NCERT), and other similar literature on test development for the selection of appropriate standardized tools. If he is developing his own tools, he should outline the procedure to be followed in their development.

A detailed discussion on sampling methods and on tools and techniques of data collecting may be found in Chapter Five.

PROCEDURES FOR TREATING DATA

In this section the researcher describes how to organize, analyse, and interpret data. The details of the statistical techniques and the rationale for using such techniques should be described in the research proposal. In case of historical research, it is necessary to throw light on the methods of internal and external criticism that are to be adhered to in the analysis of data. In documentary studies, mode of content analysis should be explained in the research proposal.

BIBLIOGRAPHY

In each research proposal, the researcher should give a list of books, journals, and other documents that he has used in selecting the problem and which he may use during the conduct of study.

TIME SCHEDULE

The researcher should also prepare a realistic time schedule for completing the study within the time available. Dividing a study into phases and assigning dates for the completion of each phase help the researcher to use his time systematically.

BUDGET SCHEDULE

The research proposals which are submitted to government, private or autonomous agencies for financial assistance should also include a budget proposal estimating the funds required for travel expenses, typing, printing and cyclostyling, purchase of equipment, tools, books and other materials.

POSTTEST

1. Describe why the researcher should be familiar with the field or area of education from which he intends to choose his problem.
2. List and describe the important sources for problem selection.
3. State what is meant by the 'definition' and 'statement' of the problem.
4. List and auswer the questions which are helpful in the evaluation of a research problem.
5. State what is meant by hypothesis.
6. Discuss the importance of hypotheses.
7. List and describe the main criteria of usable hypotheses.
8. Give some examples of usable hypotheses.
9. List and describe the important conditions conducive to the formulation of hypotheses.
10. List and describe various forms of hypotheses.
11. Give some examples of the various forms of hypotheses.
12. State how the hypotheses are tested.
13. Explain the meaning of a research proposal or synopsis.
14. Describe various types of research proposal or synopsis.
15. Describe the procedure used in the preparation of a research proposal or synopsis.
16. Outline a research proposal or synopsis in any field of education of your choice, indicating the importance of your theme.

4. Review of the Related Literature

This chapter has three sections. The first section describes the specific purposes which are served by the review of the related literature The second section provides a study guide to the researcher in locating, selecting and utilizing the primary and secondary sources of information available in the library. Finally, the third section deals with procedure which the researcher should adopt for organizing the related literature in a systematic manner.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. List and describe the specific purposes which are served by the review of related literature.
2. Describe the procedure which the researcher should adopt in locating, selecting and utilizing the primary and secondary sources of information available in the library.
3. List and state the library skills that are required by the researcher for a thorough survey of the literature related to a research topic in education.
4. Name some important reference materials and research journals which the researcher would like to consult in connection with the problem he has selected for research.
5. Describe the procedure which the researcher should adopt in organizing the related literature in a systematic manner.

Research takes the advantage of the knowledge which has accumulated in the past as a result of constant human endeavour. It can never be undertaken in isolation of the work that has already been done on the problems which are directly or indirectly related to a study proposed by a researcher. A careful review of the research journals, books, dissertations, theses and other sources of informations on the problem to be investigated is one of the important steps in the planning of any research study. As indicated in the last chapter, a review of the related literature must precede any well planned research study.

PURPOSE OF THE REVIEW

Review of the related literature, besides, to allow the researcher to acquaint himself with current knowledge in the field or area in which he is going to conduct his research, serves the following specific purposes.

1. The review of related literature enables the researcher to define the limits of his field. It helps the researcher to delimit and define his problem. To use an analogy given by Ary *et al.*, (1972, p. 56) a researcher might say:

> The work of A, B, and C has discovered this much about my question; the investigations of D have added this much to our knowledge. I propose to go beyond D's work in the following manner.

The knowledge of related literature brings the researcher up-to-date on the work which others have done and thus to state the objectives clearly and concisely.

2. By reviewing the related literature the researcher can avoid unfruitful and useless problem areas. He can select those areas in which positive findings are very likely to result and his endeavours would be likely to add to the knowledge in a meaningful way.

3. Through the review of related literature, the researcher can avoid unintentional duplication of well established findings. It is no use to replicate a study when the stability and validity of its results have been clearly established.

4. The review of related literature gives the researcher an understanding of the research methodology which refers to the way the study is to be conducted. It helps the researcher to know about the tools and instruments which proved to be useful and promising in the previous studies. The advantage of the related literature is also to provide insight into statistical methods through which validity of results is to be established.

5. The final and important specific reason for reviewing the related literature is to know about the recommendations of previous researchers for further research which they have listed in their studies.

IDENTIFICATION OF THE RELATED LITERATURE

The first step in reviewing the related literature is the identification of the material that is to be read and evaluated. The identification can be made through the use of *primary* and *secondary* sources available in the library.

In the primary sources of information the author reports his own work directly in the form of the research articles, books, monographs, dissertations or theses. Such sources provide more information about a study than can be found elsewhere. Primary sources give the researcher a basis on which to make his own judgement of the study. Though consulting such sources is a time consuming process for a researcher, yet they provide a good source of information on the research methods used.

In secondary sources, the author compiles and summarises the findings of the work done by others and gives interpretation of these findings. In them, the author usually attempts to cover all of the important studies in an area in the form of encyclopedia of education, education indexes, abstracts, bibliographies, bibliographical references and quotation sources. Working with secondary sources is not time-consuming because of the amount of reading required. The disadvantage of the secondary sources, however, is that the reader is depending upon someone else's judgments about the important and significant aspects of the study.

The decision concerning the use of primary or secondary sources depends largely on the nature of the research study proposed by the researcher. If it is a study in an area in which much research has been reported, then a review of the primary sources would be a logical first step. On the other hand, if the study is in an area in which little or no research has been conducted, a check of the secondary sources is more logical. Sources of information, whether primary or secondary, are found in a library. The researcher must, therefore, develop the expertise to use resources without much loss of time and energy. To aid the researcher in locating, selecting and utilizing the resources, a study guide is provided in relation to their use in educational research.

Locating Sources of Information

A researcher should be familiar with the library and its many facilities and services. He should also be acquainted with the regulations governing the use and circulation of materials. Many libraries use a printed guide that contains helpful information. The guide uses a diagram to indicate the location of the stacks, the periodicals section, reference section, reading rooms, and special collections of books, microfilm or microcard equipment, manuscripts, or pamphlets. The guide lists the periodicals to which the library subscribes and the names of special indexes, abstracts, and other reference materials.

The regulations concerning the use of stacks, the use of reserve books, the procedures for securing reference materials held by the library or those that may be borrowed from another library are also included in the guide.

Research scholars and other readers are usually issued a library card giving them access to the stacks. They may take the help of library staff or may carry on the independent searching for the books and other referencc materials. After using the books, it is desirable for the readers to leave them on the tables so that the library staff will return them to their proper position on the shelves.

Sometimes a reference is not available in the library. In such situation the reader must consult the 'union' catalog, which lists references found in other libraries. Such references may be obtained in the following ways by the library:

1. *By interlibrary loan system*. The reader requests the librarian to loan the desired reference from other library where it is available.

2. *By requesting a photostatic copy*. The reader may request the librarian to obtain the photostat of a page or of a number of pages of a desired reference from the source.

3. *By requesting an abstract or translation of the portion of a desired reference*. Some big libraries have abstracting and translating service that provide abstracts, or copied or translated portions of needed materials at an established fee.

4. *By requesting microfilm or microfiche*. The reader may purchase a microfilm that can be projected on library microfilm equipment. A microfiche is a sheet of film that contains micro-images of a printed manuscript or book. Its development has been one of the most significant contributions to library and information services by providing economy and convenience of storing and distribution of long runs of scholarly materials.

The size of the internationally standardized microfiche is 105 m × 148 mm which contains 98 images (7 rows with 14 micro-images) having a reduction ratio 24 x. This has been approved by the American National Micrographics Association, while the size of the British Standard microfiche is lesser which can have 60 images (BS 4187 : 1973).

A *superfiche*, which has been recently developed by the PCMI process of National Cash Register, has a 75 x reduction ratio and contains upto 1000 pages of printed material on a single 4″ × 6″ transparent card.

An even more significant development is the *ultra-fiche*. It has reduction ratio of 150 x and has the capacity of 3200 pages per fiche (Mittal, 1979, p. 10).

Various types of cameras are used to record microimage on roll film.

Planetary camera is either 35 mm or 16 mm still camera which is mounted on a vertical column that can be moved up and down to change the reduction ratio between 10x and 50x. At a time, it can be loaded with 100 feet of roll film. It does not cost much.

Step-and-repeat camera is a costly camera which is used to record automatically microimages on microfiche one by one.

Rotary camera, like the planetary camera, records microimages on roll film and has the capacity to change the reduction ratio as desired.

Flow camera costs less than half of a planetary camera. Its reduction ratio is fixed unlike previously mentioned types of camera.

All these cameras make use of *silver*, *diazo* or *vesicular* film for recording microimages.

Generally, six types of *Readers* are used for reading microfilms or microfiche (Mittal, 1979, p. 13).

(*i*) *Cuddly Microfiche Reader* is a portable reader and can be used by keeping it in one's lap. It is very cheap and can be lent to library members for home use.

(*ii*) *Micro film and Microfiche Readers.*

(*iii*) *Universal Machines.*

(*iv*) *Reader/Printer* is a push-button machine which not only helps in reading a microfilm/microfiche but is capable of producing a full-sized paper copy of the frame on the screen.

(*v*) *Production Printer/Enlarge Printer* is an automatic machine and can print the requisite number of copies of a microfilm or selected portions of a microfilm. It is used for mass production of full-sized copies of microfilms.

(*vi*) *Xerox Copyflow Machine* is a costly machine, and therefore, is beyond the reach of ordinary libraries. In can print out a microfilm into a readable size and as such a single copy of any requisite document can be had at low cost and in less time.

The Card Catalog in the Library

The card catalog is the index to the entire library. It lists the contents of publications found in the library, with the exception of serially published periodicals.

Generally, the card catalog contains author, title, and subject cards arranged alphabetically. A great deal of information about a book can be found on the cards. Besides the title of the book and the name of the author, the reader will find the date of birth of the author, the edition, the publication date, the number of pages, and the name and location of the publisher. Other items listed on the cards are bibliographies, maps, portraits, illustrations, tables, series (if any) in which a book appears, a brief description of the book whether the book is a translation and who did the translation.

Library Classification Systems

Library classification systems provide ingenious ways of systematizing the placement and location of books. Every system is based upon a methodology that is logical and orderly to the smallest detail. The two principal systems of library classification in the United States are 'Dewey Decimal' system and 'Library of Congress' system.

The 'Dewey Decimal' system is a decimal plan with the numbers running from 001 to 999.99. The 'Library of Congress' system is particularly used in large libraries. It provides for twenty main classes instead of the ten of the Dewey Decimal system. The system uses letters of alphabet for the principal headings and numerals for further sub-grouping.

In a library all books have a call number or letter that appears in the upper left-hand corner of the author, subject, or title card, and on the back of the book. These call numbers or letters are used to arrange the books serially on the library shelves and within each classification the books are arranged alphabetically by author's last name.

Library Searching Guidelines

Identifying the best available sources pertaining to a problem and extracting the essential information from them is of much importance to a researcher. For this he must develop some library searching techniques so as to save his time and effort. Van Dalen (1973, p. 88) has suggested the following worthwhile guidelines for a researcher:

1. Before using a library, familiarize yourself with its layout, facilities, services, and regulations.
2. Learn how to use the microform readers, photocopiers, and other mechanical aids.
3. In the stacks and in the periodical, reference, reserved book, and rare book rooms, note where materials that you will use frequently are placed.
4. Schedule your work session in a library when you will encounter the least competition for resources and services.
5. Make out call slips for all or most books needed in one session.
6. Copy all information that the librarian needs to obtain each reference for you, and before closing the periodical index or card catalog, recheck and rectify any errors or omissions.

7. Arrange to spend a block of time in the library that is sufficient to accomplish a specific task.
8. When little time is available, clear up questions that can be answered quickly in reference books that are readily available.
9. Before initiating search for materials in a library, write down questions that pinpoint precisely what information you wish to locate and group the questions in accordance with the areas in the library where the answers may be found.
10. Compile a list of the present and any previous names of periodicals, organizations, government agencies, research agencies collectors of statistics libraries aad museums with special collections, and outstanding authorities in your field.
11. Keep a list of the best reference books, indexes, handbooks, historical studies, and legal references in your area of specialization.
12. Obtain copies of the best bibliographies and reprint of significant research studies for your files.
13. Note which periodicals regularly or occasionally print bibliographies, reviews of the literature, or other materials and the issues in which they appear.

Reference Materials

There are a number of reference that may be useful to a researcher in the field of education. To facilitate the search for such materials, a researcher may consult the following carefully compiled volumes:

Constance M. Winchell, ed. *A Guide to Reference Books*, 8th edn. Chicago: American Library Association, 1967. This comprehensive work has biennial supplements to bring the up-to-date information in a number of languages. It describes and evaluates about 7,500 references and a section is devoted to education.

Albert J. Walford, *Guide to Reference Material.* This is a two volume work which covers (1) science and technology (1966) and (2) philosophy and psychology, religion, social sciences, geography, and history (1968).

Mary N. Barton and Marion V. Bell (1962) *Reference Books: A Brief Guide for Students and Other Users of the Library.* This guide is helpful but considerably shorter.

International Guide to Educational Documentation (1955-1960), UNESCO, 1963. This is a one-volume international guide to educational books, pamphlets, periodicals, occasional papers, film and sound recordings.

Arvid Burke and Mary Burke, *Documentation in Education.* This guide provides an excellent introduction to literature in the field of education.

The Standard Periodicals Directory. New York: Oxbridge Publishing Co., 1964-date. This is a directory of over 30,000 entries and covers every type of periodical, with the exception of local newspapers.

It is published every year and covers about two hundred classifications which are arranged by subject. An alphabetical index is provided.

Christine L. Wyner, *Guide to Reference Books for School Media Centres*. Littleton, Colo: Libraries Unlimited, 1973. This guide includes 2575 entries with evaluative comments on reference books and selection tools for use in educational institutions. It is indexed by author, subject, and title.

Encyclopaedias. This serves as a store house of information, and usually contains well-rounded discussion and selected bibliographies that are prepared by specialists. Encyclopedias are arranged alphabetically by subject, and for each field of research they present a critical evaluation and summary of the work that has been done. In addition, these suggest needed research in the field and provide *a selective bibliography*.

The following list provides a sample of encyclopedias that researchers in the field of education might use:

A Cyclopedia of Education, Paul Monroe, ed., 5 vols. (New York: Macmillan, 1911-1913). It is edited by Paul Monroe with the assistance of departmental editors and more than 1000 individual contributors. It provides excellent bibliographies and is extremely useful for historical and biographical purposes.

Encyclopedia of Modern Education, Henry D. Rivlin and H. Schueller, eds. (New York: Philosophical Library, 1943). This comprehensive work of about 200 authorities has been edited by Henry D. Rivlin and H. Schueller. It stresses present-day problems, trends, theories, and practices. The articles are accompanied by brief bibliographies and there is a system of cross references.

Encyclopedia of Educational Research. Walter Scott Monroe, ed., rev. ed. (New York: Macmillan, 1950). Monroe's *Encyclopedia of Educational Research* was prepared under the auspices of the American Educational Research Association. It aims to present a critical evaluation, synthesis, and interpretation of research studies in the field of education. All the articles, arranged alphabetically, are provided with bibliographies.

Encyclopedia of Educational Research. Chester Harris, ed.. 3d edn. (New York: Macmillan, 1960). Harris's *Encyclopedia of Educational Research* is also prepared under the auspices of the American Educational Research Association. It is not merely a revision of earlier editions, but it is completely a re-written volume that has attempted to put into a new perspective the findings of research.

Encyclopedia of Educational Research. Robert L. Ebel, ed., 4th edn. (New York: Macmillan, 1969). Ebel's *Encyclopedia of Educational Research* provides concise summaries of research and many references for further research. The articles deal with persistent educational problems and continual educational concerns.

International Encyclopedia of the Social Sciences (New York: Macmillan Co., 1968). It was prepared under the direction of ten learned societies. This reference work treats topics in all of the social sciences.

Encyclopedia of Childcare and Guidance (Garden city, New York: Doubleday and Co., 1968). It is a comprehensive treatment of the nature of the problems of childhood. It also suggests the methods of dealing with such problems.

Encyclopedia of Social Work. (New York: National Association of Social Workers, 1965). This reference work presents extensive articles on all aspects of social work.

Encyclopedia of Philosophy (New York: McGraw-Hill Book Co., 1971). This encyclopedia contains more than 7000 articles written by more than 2000 contributors in all areas of science and engineering.

Encyclopedia of Philosophy (New York: Macmillan—Free Press, 1967). It is an authoritative and comprehensive reference work covering both Western and Eastern thought—ancient, medieval, and modern.

Dictionaries. They serve as constant guides to the researcher. A few known dictionaries are given below:

Dictionary of Education. (New York: McGraw-Hill Book Co., 1973). This dictionary covers 33,000 technical and professional terms. It also includes educational terms used in various countries.

Comprehensive Dictionary of Psychological and Psycho-analytical Terms (New York: David McKay Company). It contains more than 13,000 terms. All these are defined in non-technical terms.

Dictionary of Sociology (Totowa, N.J. Littlefield, Adams and Co.). In this dictionary sociological terms are defined in non-technical language.

Roget's International Thesaurus of Words and Phrases (New York: Crowell, Collier and Macmillan). A Thesaurus is the opposite of a dictionary. One turns to the Thesaurus when one has an idea but does not yet have appropriate word to convey it. Thesaurus lists together the synonyms and antonyms of words. A researcher should use this reference in conjunction with a good dictionary to insure precision of expression.

Yearbooks, Almanacs and Handbooks. A large amount of current information on educational problems, thought and practices may be found in yearbooks, almanacs and handbooks. Some yearbooks cover a new topic of current interest each year and some give more general reviews of events. A list of some yearbooks, almanacs and handbooks is given as under:

The Handbook of Research on Teaching. N.L. Gage (ed.), (Chicago: Rand McNally & Co., 1963). This handbook presents a comprehensive research information on teaching with extensive bibliographies.

The Rand McNally Handbook of Education, Arthur W. Foshay (ed.), (Chicago: Rand McNally & Co, 1963). It is a convenient source compilation of the most important facts about education in the United States. This handbook provides a quick-reference comparison of education in England, France and USSR.

Education Yearbook (New York: Macmillan Co., 1972-date). This is an annual publication. It includes statistical data on major educational issues and movements with a comprehensive bibliography and reference guide.

Mental Measurement Yearbook (Highland Park, New Jersey: Grayphon Press, 1938-date). It is compiled by Oscar K. Buros and provides a comprehensive summary on psychological measurement and standardized tests and inventories. It is published after every four years and includes reviews on all significant books on measurement and excerpts from book reviews appearing in professional journals.

The Student Psychologist's Handbook: A Guide to Sources (Cambridge, Mass: Schenkman Publishing Co., 1969). This handbook describes the major content areas of psychology, with sources of information, methods of data collection, and the use of reference materials.

Data Processing Yearbook (Detroit: Frank H. Gille, 1952-date). This yearbook is published irregularly and includes articles on equipment, techniques, and developments in data processing. It also provides information about institutions offering data-processing and computer courses.

United Nations Statistical Yearbook. (New York: United Nations, 1949-date). This is an annual publication. It presents statistical data on population, trade, finance, communication, health, and education.

World Almanac-Book of Facts (New York: Newspaper Enterprise Association, 1968-date). This reference guide is published annually. It provides up-to-date statistics and data concerning events, progress and conditions in social, educational, political, religious, geographical, commercial, financial and economic fields.

The Standard Education Almanac. It provides a record of facts and statistics on virtually every aspect of education.

Directories and bibliographies. Directories are used by a researcher to locate the names and addresses of persons, periodicals, publishers, or organizations when he wants to obtain information, financial assistance, or research materials and equipment. Directories may help a researcher to find people or organizations who have similar professional interests or who can answer his queries or help to solve his problems.

A few examples of directories in USA and UK are:

Guide to American Educational Directories. It assembles in one volume over 12000 educational and allied directories. The directories are listed alphabetically and are arranged under subject headings.

The Education Directory (Washington: US Office of Education, Superintendent of Documents, 1912-date). This directory is published annually in five parts. It deals with names, educational agencies, officials, institutions, and other relevant data.

NEA Handbook for Local, State and National Associations (Washington, D.C.: National Education Association, 1945-date). This is an annual publication and contains listings and comprehensive reports of state and national officers of affiliated associations and departments.

Educator's World (Englewood, Colo.: Fisher Publishing Co., 1972-date). This is an annual guide to more than 1600 education associations, publications, research centres, and foundations.

National Faculty Directory. (Detroit: Gale Research Co., 1964-date). This annual publication lists alphabetically the names and addresses of more than 300,000 full-time and part-time faculty members and administrative officials of colleges and universities in the USA.

Encyclopedia of Association. (Detroit: Gale Research Co., 1964-date). This directory lists alphabetically more than 14,000 national, associations of USA. It includes information on membership, address, name of executive secretary, and statement of purpose.

Directory of Exceptional Children. (Boston: Porter Sargent Publishing Co., 1962-date). This directory provides a description of schools, camps, homes, clinics, hospitals, and services for the socially maladjusted, mentally retarded or physically handicapped in the United States.

Mental Health Directory. (Washington, D.C.: National Institute of Mental Health, Government Printing Office, 1964-date). This annual publication lists national, state and local mental health agencies in the USA.

Americam Library Directory. (New York: R.R. Bowker Co., 1923-date). This directory provides a biannual guide to private, state, municipal, institutional, and collegiate libraries in the USA and Canada. It includes information on special collections, number of holdings, staff salaries, budgets and affiliations.

Kelley, Thomas. (ed.) *A Select Bibliographies of Adult Education in Great Britain* (London: National Institute of Education, 1952). Blackwell, A.M. *A list of Researches in Educational Psychology presented for Higher Degrees in the Universities of the United Kingdom and the Irish Republic from* 1918. (London: Newnes Educational Publishing Co., 1950).

In India, a very few bibliographical guides to educational research on a national basis have appeared. *Bibliography of Doctorate Theses in Science and Arts* accepted by the Indian Universities for 1946-48 and 1948-50 has been published recently by the Inter-university Board of India. Theses are listed under university, with subject sub-headings including education.

The index. A periodical index serves the same purpose as the index of a book or the card file of library. It identifies the source of the article or of the book cited by listing the titles alphabetically, under author and the readers should read all such directions before trying to locate the references.

A list of some important educational indexes is given below:

Edueation Index. (New York: H.W. Wilson Co., 1929-date). One valuable and work saving guide created for educators is *Education Index*. It is published monthly (September through June), cumulated annually and again every three years. It indexes more than 250 educational periodicals, and many yearbooks, bulletins, and monographs published in the United States, Canada, and Great Britain. The materials on adult education, business education, curriculum, educational administration, educational psychology, educational research, exceptional children, higher education, guidance, health and physical education, international education, religious education, secondary education and teacher education are included in this index.

Canadian Education Index (Ottawa, Ontario: Canadian Council for Educational Research, 1965-date). This index is issued quarterly and indexes periodicals, books, pamphlets, and reports published in Canada.

Current Index to Journals in Education (New York: Macmillan Information, 1969-date). This index is published monthly and cumulated six monthly and annually. It indexes about 20,000 articles each year from more than 700 education and education-related journals, under author and subject headings.

ERIC Educational Documents Index. (Washington, D.C.: National Institute of Education, Government Printing Office, 1966-date). This index is published annually. It is a guide to all research documents in the Educational Resources Information Centre collection.

Index of Doctoral Dissertations International. (Ann Arbor, Mich.: Xerox University Microfilms, 1956-date). Published as the issue 13 of *Dissertation Abstracts International* each year, it consolidates into one list all dissertations accepted by American, Canadian, and some European universities during the academic year, as well as those available in microfilm.

International Guide to Educational Documentation. (Paris: UNESCO). This guide is published every five years. It indexes annotated bibliographies covering major publications, bibliographies and national directories written in English, French and Spanish.

British Education Index Vol I, August, 1954 to November, 1958. This index is compiled by the Librarians of Institutes of Education and it includes references to articles of educational interest which have been published in more than fifty periodicals during the period of four years.

Index to Selected British Educational Periodicals (Leeds: Librarians of Institutes of Education, 1945-date). This index is issued thrice per year and it covers 41 educational periodicals excluding the periodicals on fundamental and adult education.

Research Periodicals

Information about new ideas and developments often appear in periodicals long before it appears in books. There are many periodicals in education and in other closely-related areas that are the best source for reports on recent research studies. Such periodicals give much more up-to-date treatment to current questions in education than books possibly can. They also publish articles of temporary, local or limited interest that never appear in book form. The periodicals of proper dates are the best sources for determining contemporary opinion and status, present or past.

It has been estimated that there are about 2100 journals that are specifically related to the field of education. In all such journals one may also find articles of interest devoted to psychology, philosophy, sociology, and other subjects.

All those engaged in educational research should become acquainted with certain educational periodicals and they should also learn to use the indexes to them. Knowledge about the editor of a periodical, the names, of its contributors, and the associations or institutions publishing it may serve as clues in judging the merit of the periodical.

Ulrich's Periodicals Directory; *A Classified Guide to a Selected List of Current Periodicals, Foreign and Domestic*, 12th ed. (New York: Bowker, 1967) provides a comprehensive list of periodicals relating to education. In this directory, periodicals are grouped in a subject classification and are alphabetically arranged. Each entry includes title; sub-title, date of origin, frequency of publication, annual index, cumulative indexes, and item characteristics of each periodical.

In India, many periodicals are published by some associations or institutions. They provide a medium for dissemination of educational research and exchange of experience among research workers, teachers scholars and others interested in educational research and related fields and professions.

Abstracts

Abstracts include brief summaries of the contents of the research study or article. They serve as one of the most useful refernce guides to the researcher and keep him abreast of the work being done in his own field and also in the related fields.

In America the most useful of these references are the following:

The review of educational research. It gives the educator an excellent overview of the work that has been done in the field and keep him well informed about the recent developments. This publication, between 1931 and 1969, reviewed about every three years each of the following eleven major areas of education: (*i*) Administration; (*ii*) Curriculum; (*iii*) Educational Measurement; (*iv*) Educational Psychology; (*v*) Educational Sociology; (*vi*) Guidance and Counselling; (*vii*) Language Arts, Fine Arts, Natural Sciences, and Mathematics; (*viii*) Research Methods; (*ix*) Special Programmes; (*x*) Mental and Physical Development; and (*xi*) Teacher Personnel.

Since June, 1970, the *Review of Educational Research* has pursued a policy of publishing unsolicited reviews and research on topics of the contributor's chosing. The role played by this publication in the past has been assumed by the *Annual Review of Educational Research*. *Research in Education* (*RIE*). This represents the most comprehensive publication of research materials in education today. RIE is published monthly since 1966 by the Educational Resources Information Centre (ERIC) and indexed annually. Each monthly issue of RIE is divided into three sections: (1) Document Section; (2) Project Section; and (3) Accession Numbers Section.

Psychological abstracts. This useful reference is published by the American Psychological Association since 1927. It is published bimonthly and contains abstracts of articles appearing in over 530 journals, mostly educational periodicals. The biannual issues (January-June, July-December) contain both an author and subject index.

Education abstracts. This is a publication of UNESCO, which began in 1949 and has been published monthly except in July and August. Each introductory essay devoted to a particular aspect of education, is followed by abstract of books and documents, selected from various countries dealing with the topic under consideration.

In addition to the above periodicals, a researcher may also consult the following publications:

(1) *Annual Review of Psychology* (1950-date):
(2) *Child Development Abstracts and Bibliography* (1927-date);
(3) *Psychological Bulletin* (1904-date);
(4) *Sociological Abstracts* (1952-date);
(5) *Educational Administration Abstracts* (1966-date);
(6) *Sociology of Education Abstracts* (1965-date);
(7) *Mental Retardation Abstracts* (1964-date);
(8) *Dissertation Abstracts International* (1952-date).

In India, the Ministry of Education and Social Welfare, Government of India, New Delhi, since 1955, has been issuing *Indian Education Abstracts* every three months. It abstracts the contents of books and periodicals on education published in India in English and Hindi. Most of the Indian Journals are indexed in the various issuss. The abstracts appear under the following subject headings: Philosophy of Education; Educational Psychology; Measurement and Testing; Examinations; Students and Students' Organizations; Educational and Vocational Guidance; Teacher Education and Training; Curriculum; Basic Education; Health and Physical Education; Primary, Secondary, Higher, Vocational and Technical Education; and Social Education.

Many professional periodicals and yearbooks, in India and abroad include some reviews of research and technical discussions of educational problems in one or all of their series. A list of some of the publications are:

USA; *Journal of Educational Research, NEA Research Bulletin, Educational and Psychological Measurement, Journal of Experimental Education, Research Quarterly, Journal of Research in Music Education,* *American Educational Research Journal, Reading Research Quarterly, Journal of Educational Psychology, Journal of Psychology, Journal of Social Psychology, Journal of Applied Psychology, Sociology of Education, American Journal of Sociology, American Sociological Review, Sociology and Social Research, Harvard Educational Review, Journal of Teacher Education, Elementary School Journal, History of Education Quarterly, and Educational Forum.*

UK: British Journal of Educational Psychology.
India: Indian Educational Review;
Journal of Psychological Researches:
Indian Journal of Applied Psychology;
Indian Journal of Experimental Psychology;
Journal of Education and Psychology;
The Education Quarterly.

Theses and Dissertations

Theses and dissertations are usually preserved by the Universities that award the authors their doctoral and masters degrees. Sometimes these studies are published in whole or in part in various educational periodicals or journals. Because the reports of many research studies are never published, a check of the annual list of theses and dissertations issued by various agencies is necessary for a through coverage of the research literature.

In the United States, doctoral dissertations in all fields, including education, can be found in sources compiled by various agencies. For the period 1912 to 1938, the Library of Congress issued the annual *List of American Doctoral Dissertations* for published studies. The Association of Research Libraries published the list of *Doctoral Dissertations Accepted by American Universities* from 1933-1934 to 1954-1955. This service was continued by the Index to *American Doctoral Dissertations* 1956-1963, which became the *American Doctoral Dissertations*, 1963-64 to date. It lists all doctoral dissertations accepted by the American and Canadian universities and other educational institutions.

Dissertation Abstracts International, May, 1970, abstracts dissertations in the humanities, social sciences, physical sciences and engineering. It is published monthly. For each dissertation there is a 600 word abstract that provides the researcher enough information to satisfy his needs. If a researcher wants to read a complete study of a dissertation that is presented in *Dissertation Abstracts International*, he can purchase a microfilm or xerox copy from University Microfilms. The order under and price are provided in the abstract.

In India, most universities publish abstracts of dissertations and theses that have been completed at the institution.

Kurukshetra University, Kurukshetra (Haryana) published *Abstract of M. Ed. Dissertations* Vol. I, 1966; *Abstracts of M.Ed Dissertations*, Vol. II, 1967; *Abstract of M.Ed Dissertations*, Vol. III, 1968; *Abstract of M.Ed Dissertations* Vol. IV, 1969; *Abstract of M.Ed Dissertations*, Vol. V, 1970, *Abstract of M.Ed dissertations and Ph.D. Theses*. Vol. VI, 1973.

M.B. Buch ed. *A Survey of Research in Education*, Centre of Advanced Study in Education, Baroda: M.S. University, 1973. This publication contains all the research studies in education completed in Indian universities upto 1972. The break up of the studies in the said volume is 462 Ph.D. studies and 269 project research. The abstracts of all the studies have been classified into seventeen meaningful educational areas. They are (*i*) Philosophy of Education, (*ii*) History of Education, (*iii*) Sociology of Education, (*iv*) Economics of Education. (*v*) Comparative Education, (*vi*) Personality, Learning and Motivation, (*vii* Guidance and Counselling, (*viii*) Tests and Measurement, (*ix*) Curriculum, Methods, and Textbooks; (*x*) Educational Technology, (*xi*} Correlates of Achievement, (*xii*) Educational Evaluation and Examination, (*xiii*) Teaching and Teaching Behaviour, (*xiv*) Teacher Education, (*xv*) Educational Administration, (*xvi*) Higher Education, and (*xvii*) Non-formal Education.

M.B. Buch, ed. *Second Survey of Research in Education (1972-1978)* Baroda: Society for Educational Research and Development (1979). This publication incorporates 839 research studies completed during the period 1972-1978 and follows the same pattern of organization of seventeen research areas as *A Survey of Research in Education*. The first chapter gives a broad perspective of the place and function of research for educational development including historical account of the development of educational research in India. Each subsequent chapter includes a report based on the abstracts on research studies giving the trend of research in the area, including the gaps and highlighting the research priorities as perceived by the author. The abstracts are arranged alphabetically for each area and continuously numbered throughout the book. Each abstract contains the title of the study, the objectives and/or hypotheses examined, methodology including the sample, tools of research, the statistical techniques used, and the findings. A special feature of this publication is the incorporation of a large number of studies on educational problems completed in the university departments of social sciences and humanities other than the departments of education. The trend reports are based not on the research completed during the period 1972-1978, but on the total research activities during the period 1940 to 1978.

Newspapers and Pamphlets

Many articles of particular interest to a researcher may be located through pamphlets and newspapers. Current newspapers provide up-to-date information on speeches, seminars, conferences, new trends, and a number of other topics. Old newspapers, which preserve a record of past events, movements and ideas, are particularly useful in historical inquiries. Some libraries catalog pamphlets and newspapers in their reference sections.

Government Documents

Government documents are a rich source of information. They include statistical data, research studies, official reports, laws and other materials that are not always available elswhere. These are available in national, regional, state as well as local level government offices.

Monographs

Monographs are also major research sources. In the United States, universities and teachers' colleges publish many research studies in education in the form of monographs. A few examples of these are *Supplementary Educational Monographs*, *Educational Research Monographs*, and *Lincoln School Monographs*. In England too, the various institutes of education publish monographs from time to time. In India, only a limited number of monographs are published by some universities and research organizations.

Computer-generated Reference Materials

School Research Information Service (SRIS), Direct Access to Reference Information (DATRIX), and Psychological Abstract Search and Retrieval Service (PASAR) in the United States provide a number of computer-generated reference sources that may save a great deal of time and effort of the researcher. SRIS operated by Phi Delta Kappa (Bloomington, Indiana) provides a computer printout of abstracts for a moderate fee. DATRIX, a development of the University Microfilms (Ann Arbor, Michigan) provides computerized retrieval for *Dissertation Abstracts*, from 1928 to date. The researcher can procure information on microfiche or Xerogrephic copy of the complete dissertation which he needs from *Universitiy Microfilms* on payment. The PASAR furnishes printouts of abstracts of psychological journal articles, monographs, reports, and parts of books for a moderate fee.

ORGANIZING THE RELATED LITERATURE

After making the comprehensive survey of the retated literature, the next step for the researcher is to organize the pertinent information in a systematic manner. It should be done in such a way as to justify carrying out the study by showing what is known and what remains to be investigated in the topic of concern. According to Ary *et al.* (1972, p. 67):

> The hypotheses provide a framework for organizing the related literature. Like an explorer proposing an expedition, one maps out the known territory and points the way to the unknown territory he proposes to explore. If the study has several aspects, or is investigating more than a single hypothesis this is done separately for each facet of the study.
>
> One should avoid the temptation to present the literature as a series of abstracts. Rather, it should be presented in such a way as to lay a systematic foundation for the study.

The organization of the related literature involves recording the essential reference material and arranging it according to the proposed outline of the study.

Recording Reference Information

Once pertinent information has been identified, the researcher should record certain essential information for locating the material on 3 × 5 inch index card to serve as a *bibliography card.* To make writing of the final report simpler, it is desirable that the information recorded in the bibliography should appear, in content and style, exactly as it will appear in the final report.

The basic information in the bibliography card should include name of the author with last name first; title of the book or article; name of the publication (for articles); name of the publisher; date of publication; volume number; page numbers; and library call number (for books). If some of this information is not available the specified space should be left blank so that the missing information can be included immediately upon locating the references.

Recording Content of References

After recording the essential information on the bibliography cards, it is necessary to arrange the cards according to the location of the material in the library. For example, the researcher may list together all cards pertaining to the materials located in the periodical section. Similarly, all the materials located in the reserve section may constitute another list, and so on. Then the researcher should make a systematic review of the material located in a specific section of the library and after reviewing each reference on the list he should proceed to another list.

All the information likely to be used in the final report should be recorded on a 4 × 6 inch card to serve as a *content card.* The information to be recorded on the content card will depend on the source from which it is taken. If it is from a primary source, it may include brief bibliographic information comprising of author's last name, brief title of the report, specific page numbers on which information is located; sentence statement of the problem; brief description of the study; statements of findings or conclusions, or both; a card code as to the aspect of the research to which the material most closely relates.

The information to be recorded from the secondary source is somewhat different from the primary source. Turney and Robb (1971, p. 55) have given the following suggestions for recording information from a secondary source:

1. Provide brief bibliographic information (as with a primary source).
2. Record on a single card only those statements that are related to the same topic (if all the information cannot be placed on one card continue statements on another card and staple to the first card).
3. Paraphrase, in complete statements, the most relevant ideas. Record direct quotations only if they are stated concisely and effectively, and if paraphrasing might change the meaning.
4. Place a page number and a paragraph number after each separate statement indicating its location in the reference in case you need to review it again.
5. Code the cards (probably in the upper-right hand corner) according to topic (s) to which it most closely relates.

Preparation of the Related Literature Report

For the preparation of the report of the related literature, the researcher should arrange the bibliographic and content cards according to the proposed outline of the problem. This can be done with the help of card code.

The report of the related literature should begin with an introductory paragraph describing the organization of the report. After the introduction, the researcher should present the studies most relevant to each aspect of the proposed problem outline. Studies with similar and contradictory results should be reported side by side without using excessive space.

POSTTEST

1. What is the importance of survey of related literature in educational research? Illustrate by taking a specific research problem as to how the survey of the related literature can be helpful at various stages.
2. Describe the procedure which the researcher should adopt in locating selecting and utilizing the primary and secondary sources of information available in the library.
3. What library skills are required for a thorough survey of literature related to a research topic in education?
4. Name some important reference books with authors' names and some important educational journals you would like to consult in connection with the problem you have selected for research.
5. Describe the procedure which the researcher should adopt in organizing the related literature in a systematic manner.

5. Data Collection

This chapter has two major sections. The first section deals with the definitions of population and sample and the steps involved in the process of sampling. It also describes various methods of sampling: probability and non-probability sampling. The second section deals with the development and construction of various data gathering tools: psychological tests, inquiry forms, observation, interview and sociometric methods. It also describes the advantages and limitations of these tools.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Define the terms population and sample.
2. List and describe the steps involved in the process of sampling.
3. List the various methods of sampling.
4. Define a non-probability sample,
5. List and describe the various types of non-probability sample.
6. Give some examples of the various types of non-probability sample.
7. Difine a probability sample.
8. List and describe the various types of probability sample.
9. Give some examples of the various types of probability sample.
10. Define the terms statistics and parameter.
11. List and describe the various sampling errors.
12. List the various data gathering tools.
13. Describe the nature and types of psychological tests.
14. List and describe the evaluative criteria of a psychological test.
15. Desfiine the terms validity, reliability and usability in the context of psychological tests.
16. List and describe the various types of validity and reliability.
17. Describe the general procedure used in the development of a psychological test.
18. Name some important tests of intelligence, aptitude, creativity, and personality developed or adapted in India.
19. Describe the nature and various types of inquiry forms.
20. Define a questionnaire.
21. List and describe the various types of questionnaires with their examples.
22. List and describe the steps involved in the construction of a questionnaire.
23. Describe the uses and limitations of a questionnaire.
24. Define a schedule.
25. List and describe the steps involved in the construction of a schedule.

26. Define a checklist.
27. List and describe the steps involved in the construction of a schedule.
28. Describe the uses of (*i*) a schedule and (*ii*) a checklist.
29. Define a rating scale.
30. List and describe the various types of rating scale.
31. List and describe the steps involved in the construction of a rating scale.
32. Describe the advantages and limitations of rating scales.
33. Define a score card.
34. Describe the advantages and limitations of a score card.
35. Define the term attitude.
36. List and describe some important scaling techniques.
37. Describe the procedure involved in the construction of an attitude scale following the methods of equal appearing intervals and summated ratings.
38. List and describe the uses and limitations of attitude scales.
39. Define observation
40 List and describe the steps involved in the process of observation.
41. Describe the advantages and limitations of observation.
42. Define interview.
43. List and describe the various types of interview.
44. List and describe the steps involved in the process of interview.
45. Describe the advantages and limitations of an interview.
46. Define sociometry.
47. List and describe various sociometric techniques.
48. Describe the advantages and limitations of sociometric techniques.

Data collection is essentially an important part of the research process so that the inferences, hypotheses or generalizations tentatively held may be identified as valid, verified as correct, or rejected as untenable. In order to collect the requisite data for any research problem, the researcher has to sample the population concerned, since it is not possible to encompass the entire population, to devise appropriate tools for measuring the attributes concerned and to administer these tools.

SAMPLING

Most of the educational phenomena consist of a large number of units. It would be impracticable, if not impossible, to test, to interview or observe each unit of the population under controlled conditions in order to arrive at principles having universal validity. Some populations are so large that their study would be expensive in terms of time, money, effort, and manpower. Sampling is the process by which a relatively small number of individuals or measures of individuals, objects, or events is selected and analysed in order to find out something about the entire population from which it was selected.

It is often desirable in order to reduce expenditure, save time and energy, permit measurement of greater scope, or produce greater precision and accuracy. Sampling procedures provide generalizations on the basis of a relatively small proportion of the population.

(The representative proportion of the population is called a *sample.*) To obtain a representative sample, the researcher selects each unit in a specified way under controlled conditions. Usually four steps are involved in the process:

1. Defining the population.
2. Listing the population.
3. Selecting a representative sample.
4. Obtaining an adequate sample.

Defining the Population

A population refers to any collection of specified group of human beings or of non-human entities such as objects, educational institutions, time units, geographical areas, prices of wheat or salaries drawn by individuals. Some statisticians call it universe.

A population containing a finite number of individuals, members or units is called a finite population. Such, for instance, is the population of weights of thirty boys in a class. A population with infinite number of members is known as infinite population. The population of pressures at various points in the atmosphere is an example of infinite population. The population of concrete individuals is called as existent population, while the collection of all possible ways in which an event can materialize, as the hypothetical population. All the 400 students of tenth grade of a particular high school is an example of existent type of population and the population of heads and tails obtained by tossing a coin an infinite number of times is an example of hypothetical population.

The population is properly defined so that there is no ambiguity as to whether a given unit belongs to the population. For example, in a survey of achievement in mathematics a researcher will have to define the population of students by age or by grade and, if necessary, he will also specify the type of schools, the geographical area and the academic year for which the data will be collected. Inferences concerning a population cannot be drawn until the nature of the units that comprise it is clearly identified. If a population is not properly defined, a researcher does not know what units to consider when selecting the sample.

Listing the Population

In order to select a sample from a given population it is also necessary to have a complete, accurate, and up-to-date list of all the units in the population. Such a list is known as a *sampling frame*. If a sampling frame is not readily available at the time of investigation, it is essential to prepare a frame before the samples can be drawn. This task is time consuming and sometimes obstacles may arise that will prevent the researcher from obtaining the required information. It is very difficult to prepare a perfect sample frame especially in large scale surveys. The following are some possible defects of such frames:

1. The frame may be inaccurate when it either contains wrong information about the units or it lists the units which do not exist.
2. The frame may be incomplete when all the units are not listed.
3. The frame may contain units which are listed twice.
4. The frame may be inadequate when all the categories of material of the target population are not included.
5. The frame may be out of date when it contains old frames.

Selecting a Representative Sample

After defining a population and listing all the units, a researcher selects a sample of units from the sampling frame. The process of such a selection is called sampling. To serve a useful purpose sampling should be unbiased or representative. A good sample must be as nearly representative of the entire population as possible and ideally it must provide the whole of the information about the population from which the sample has been drawn. The logic of the theory of sampling is the logic of induction, that is we proceed from particular, (i.e. sample) to general (i.e. population) and all the results are expressed in terms of probability.

Obtaining an Adequate Sample

If the population under study is homogeneous, a small sample is sufficient. On the other hand, a much larger sample is necessary if there is greater variability in the units of the population. But increasing the size of the sample is of little value, if units are not chosen in a way that ensures representativeness of the sample. A small sample is often satisfactory in an intensive laboratory in which greater precision is desired and in which many measurements are taken upon each subject. Garrett (1962, p. 208) has pointed out that if the size of the sample is less than 25, there is often little reason for believing such a small group of units to be adequately descriptive of any population. If a greater precision of results in an investigation is needed, the sample should be larger. Greater precision is sometimes needed when one is dealing in areas where differences are likely to be small. If the differences between the variables under study are large or when variables are highly correlated, relatively small samples can be used because greater precision is not needed. If the differences are small, a greater precision is needed to locate them, and so larger samples are needed. The adequacy of a sample will depend upon our knowledge of the population as well as upon the method used in drawing the sample.

METHODS OF SAMPLING

Sampling methods can be classified into two broad categories:

1. Non-probability sampling.
2. Probability sampling.

Non-Probability Sampling

(In non-probability sampling, the units are selected at the discretion of the researcher.) Such samples derive their control from the judgement of the researcher. If he is to select a sample of 200 school students, he may exercise his own judgement based on experience for including a given student in the sample. The sample so selected is called a *judgement* or *purposive* sample. (Such a sample is arbitrarily selected because there is good evidence that it is a representative of the total population. The evidence is always based on experience. (For example, in a study of delinquency, the researcher, on the basis of his past experience, selects certain areas that have shown higher rate of delinquent acts in the past.)

Suppose a researcher wants to study the attitude of teachers towards co-education at the secondary school stage. He may fix a quota of 300 teachers and go on collecting data from first 300 teachers he is able to contact and gets cooperation from them. The sample of 300 teachers thus selected is called quota sample. If the quota is not fixed and the researcher collects information from all the teachers that are conveniently available, then the sample is called incidental sample. Educational researchers, because of administrative difficulties in applying probability sampling methods, often use available groups as samples.

The non-probability sampling methods are very convenient in the situations when the sample to be selected is very small and the researcher wants to get some idea of the population characteristics in a short time. Such samples are used in the situations where the researeher does not want a representative sample but to gain insight into the problem by selecting only informed persons who can provide him the maximum degree of insight into his problem with comprehensive information.

The following are some limitations in the use of non-probability sampling methods:

1. No confidence can be placed in the data obtained from such samples and results thus obtained cannot be generalized for the entire population. Generalizations beyond any sample can be made safely only when it represents the population in every significant respect.

2. The selective sampling based on convenience affects the variance within groups as well as between groups and so sampling error of such samples cannot be determined.

3. Non-probability samples depend exclusively on uncontrolled factors and researcher's insight, and there is no statistical method to determine the margin of sampling errors.

4. Sometimes such samples are based on an obsolete frame which does not adequately cover the population.

Probability Sampling

In probability sampling the units of the population are not selected at the discretion of the researcher but by means of certain procedures which ensure that every unit of a population has one fixed probability of being included in the sample. Such a method is also called random sampling.

According to Good (1966, p. 217) probability samples have the following characteristics;

1. Each unit in the sample has some known probability of entering the sample.

2. Weights appropriate to the probabilities are used in the analysis of the sample.

3. The process of sampling is automatic in one or more steps of selection of units in the sample.

There are different methods of selecting a probability sample, all of which have their own advantages and limitations.

1. *Simple or unrestricted random sampling.* In simple or unrestricted random sampling each unit of the population is given an equal chance of being selected. The selection of units from the population is done in such a manner that every unit in the population has an equal chance of being chosen, and the selection of any one unit is in no way tied to the selection of any other. The law of chance is allowed to operate freely in the selection of such a sample and carefully con-

trolled conditions are created to ensure that each unit in the population has an equal chance of being included in the sample. To prevent the researcher from biasing the results by exercising direct control over the selection of units, several devices are employed to draw samples from populations.

(A researcher may use lottery method to draw a simple random sample.) In that case he will use a set of N tickets, if there are N units in the population, with numbers 1 to N written on these tickets. The tickets will be thoroughly mixed up and then N tickets will be drawn one by one, if the sample of size N is required. The units which have the serial numbers occuring on these N tickets will be considered as selected. If a small group is involved, a coin may be flipped to select the sample by assigning head or tail to an unit.

The best method of drawing a simple random sample is to employ a Table of Random Numbers, such as those prepared by Fisher and Yates, Tippet, or Kendall and Babington-Smith. After assigning consecutive numbers to units of the population, the investigator starts at any point on the table of random numbers and reads consecutive numbers in any direction, horizontally, vertically, or diagonally. When a number is read that corresponds with that written on a unit card, that unit is chosen for the sample.

As an illustration suppose a sample of 12 schools is to be selected at random from a serially numbered population of 230 schools. Using a portion of a table of random numbers reproduced here, 12 three digit numbers (because total population of schools is a three-digit figure of 230) are selected by reading Table 5.1.

TABLE 5.1. An Abbreviated Table of Random Numbers

Row	*Col*	(*1*)	(*2*)	(*3*)	(*4*)	...	(*14*)
1		10480	15011	01536	02011	...	90700
2		22368	46573	25595	85393	...	99505
3		24130	48360	22527	97265	...	58629
4		42167	93093	06243	61680	...	16379
5		37570	39975	81837	16656	...	54613
6		77921	06907	11008	42751	...	42880
7		99562	72905	56420	69994	...	12952
8		96301	91977	05463	07972	...	32307
...		...	14342	...	...	...	...
...		...	36857	...	...	...	...
...		...	69578	...	...	...	...
...		...	...	...	...	...	...
...		...	...	...	...	...	...
...		...	...	...	...	...	...
...		...	...	...	...	...	...
...		...	...	...	...	...	...
...		...	...	...	...	...	...
60		78919	19474	23632	27889	...	87820

If the researcher starts with the first row and the first column, 104 is the first three-digit number, 223 is the next number and so on. He can start with any row and any column, and the number occurring in it as the first number to start with. Suppose the researcher selects fourth row and second column, the number so fixed is 930. He, therefore, can select the first 12 numbers from this column starting with 930.

The Sample

~~930~~	~~368~~	127✓	~~584~~	~~575~~	~~579~~	~~627~~	~~522~~	~~819~~
~~399~~	~~695~~	213✓	~~323~~	042✓	~~834~~	075✓	~~674~~	~~351~~
069✓	~~409~~	~~540~~	~~270~~	~~264~~	~~425~~	~~659~~	003✓	167✓
~~729~~	~~930~~	~~539~~	~~330~~	047✓	~~563~~	~~282~~	~~683~~	~~839~~
~~919~~	~~611~~	~~976~~	~~722~~	~~698~~	185✓	~~549~~	104✓	~~350~~
143✓	~~973~~	~~912~~	205✓	~~657~~	~~896~~	~~295~~	~~818~~	~~047~~

(In selecting this sample of 12 schools, forty two numbers were deleted because they were larger than the population of 230 described. The last number 047 was deleted because it duplicated previous selection).

The numbers so selected are 069, 143, 127, 213, 205, 042, 047, 185, 075, 003, 104, and 167. Sometimes to avoid the larger deletions, the numbers greater than the total of 230 may be replaced by the remainders which are got by dividing the numbers by 230. One caution is, however, necessary. Since the largest three digit number which is a multiple of 230 is 920, we should not select any number which is greater than 920. Thus the numbers 930, 939, 973, etc. will not be considered in the replacement process. The first number 399 can be replaced by $169=(399-230\times 1)$. Similarly, 729 can be replaced by $39=(729-230\times 3)$ and so the numbers 399, 729, 919, 368, 695, 409, 611, and 540 will be replaced by 169, 39, 229, 138, 5, 179, 151 and 80. Finally, units with serial numbers 169, 69, 39, 229, 143, 138, 5, 179, 151, 127, 213 and 80 will be selected to get a simple random sample of 12 schools. However, if any number is repeated in the table it should be substituted by another next number from the same column.

For populations to 99 in number, two digits are selected, from 100-999, three digits, and from 1000 to 9999, four digits. If one column is exhausted, the researcher may go on to the next column until a sample of the desired size is obtained. In some situations before the desired sample is obtained, some selected units may decline to participate in the investigation or during the process of investigation, some units may dropout. A sampling bias will be introduced, if these units differ from the remaining units in characteristics that are known to be related to the phenomenon under investigation. It is, therefore, necessary for a researcher to report not only the size of the sample but also the number of dropouts and the number of the selected units that did not participate in the investigation. The variables relevant to the study in which the dropouts and non-participants differed from the remaining participants may also be mentioned by the researcher.

The simple random sampling, which ensures the best results, in an ideal plan theoretically. However, from a practical point of view, a complete listing of population, unit by unit may not be possible. If possible, it may involve very high cost which a researcher or an organization may not be able to afford.

Simple random sampling is neither possible nor feasible if lists of units do not exist or if such lists are incomplete. Sometimes lists are available, but not in such a form as to permit simple random sampling. Moreover, if there is more heterogeneity among the units of the population a simple random sample may not necessarily represent the characteristics of the total population, even if all selected units participate in the investigation.

Stratified Random Sampling

At times a simple random sample, particularly a small one, may by chance have an undue proportion of one type of units in it and, therefore, it is necessary to make certain that the units included in the sample are selected in proportion to their occurrence in the population. When the units in a sample are proportional to their presence in the population, the sample is said to be stratified. When employing the method of stratified random sampling in a researcher divides his population into different strata by some characteristic which is known from previous research or theories to be related to the phenomenon under investigation, and from each of the smaller homogeneous groups falling in each strata he draws randomly a predetermined number of units. Thus, in addition to randomness, stratification introduces a secondary element of control as a means of increasing precision and representativeness.

The usual startification factors are sex, age, socio-economic status, educational background, residence (urban or rural), occupation, religion, caste, general intelligence and so on. The efficiency of the stratified random sample depends upon the allocation of sample size to strata. The simplest and most common system of allocation of sample units among strata is in proportion to the size of the strata.

If the number of units selected from different strata are proportional to the total number of units in the strata, the sample is said to be selected with *proportional allocation*. For example, a District Education Officer has a total population of 1000 teachers serving in different schools of his district and he wishes to determine his teacher's reactions to new educational pattern of 10+2+3 using a questionnaire. The researcher wants to analyse the responses of teachers according to their ages, because their reactions may be influenced by the characteristics of age. For this he may categories the population of teachers in various groups as given in the Table 5.2 below:

TABLE 5.2. Distribution of Teachers According to Their Ages

Age in years	*Frequencies*
21—27	120
28—34	180
35—41	300
42—48	220
49—55	180
	N=1,000

The researcher has stratified the population of 1000 teachers into five strata on the characteristic of age. Suppose he wants to select a 25 per cent sample with proportional allocation from the population. To ensure that the sample is stratified on age, the researcher will need to take 25 per cent of the teachers in each of the five stratum. The number of teachers in the population falling in the 21-27 years stratum is 120 and so he must select 30 teachers from his stratum. Similarly to get 25 per cent teachers from other four strata, the researcher needs to select 45, 75, 55 and 45 from 28-34, 35-41, 42-48 and 49-55 age groups respectively. To ensure further that the sample is representative, the units (teachers) within each stratum must be selected randomly.

It is also possible to stratify a population on more than one characteristic within the same study. For example, the principal of a college wishes to determine the opinion of its students regarding the functioning of the college cafeteria. He feels that the opinions of boys and girls may differ, and also may differ among P.U.C.; T.D.C. Part I; T.D.C. Part II and T.D.C. Part III students. Suppose the population of the college by age and class levels is as given in the Table 5.3.

TABLE 5.3. Classwise Distribution of Boys and Girls.

	Boys	*Girls*	*Total*
P.U.C.	120	80	200
T.D.C. Part I	100	52	152
T.D.C. Part II	140	56	196
T.D.C. Part III	88	44	132
Total	448	232	680

The total population of the college consists of 680 students. In order to ensure proportional representation of boys and girls within each class, the principal should select a stratified sample from the population of 680 students. The principal wishes to select a stratified sample of 170 students which represents 25 per cent of the population which is divided into eight strata. To obtain a stratified sample, he must take 25 per cent of the students from each of the eight stratum. For this he will select randomly 30 boys and 20 girls from P.U.C.; 25 boys and 13 girls from T.D.C.-I, 35 boys and 14 girls from T.D.C.-II; and 22 boys and 11 girls from T.D.C.-III.

It may be noted that when a population is stratified, the units within each stratum are more or less homogeneous than are the units within the entire population. For example, suppose a population is stratified into three strata on the basis of intelligence—high, middle, and low intelligence. It is evident that the variability of intelligence in the total population is greater than the variability within any particular stratum. Also, in cases where two characteristics are perfectly correlated in the population, stratification on one characteristic would lead to perfect stratification on the other characteristics. On the other hand, when two characteristics are only moderately correlated, stratification on one characteristic will not lead to perfect stratification. However, in such a case the stratification on one characteristic will have the effect of reducing the heterogeneity of the second characteristic within each stratum. When a researcher stratifies a population on one characteristic, he should expect the population to vary on other characteristics. Stratified random sample is very useful when lists of units or individuals in the population are not available. This method has been found practical even for small finite populations when cent per cent response is difficult to secure in the desired time. Stratified random sampling provides more accurate results than simple random sampling only if stratification results in greater homogeneity within the strata, with respect to trait under study than it would be found in the whole population taken as a unit. Stratification is particularly useful in opinion survey studies in which, for example, the categories of 'strongly approve', 'approve', 'undecided', 'disapprove' and 'strongly disapprove' might be related to such factors as sex, educational status, socio-economic status, and so on.

Systematic Sampling

When a frame of a population is available or when a population can be accurately listed and is finite, a method of systematic selection will provide a sample which approximates a random sample. In systematic sampling, a researcher generally starts with a list in which all the N units of the population are listed in alphabetic or some other order. To select a sample of size n, he has to select a unit at random from the first k units of the list and then every kth subsequent unit is selected. The number k is so chosen that nk is equal to N or is just less than N. To illustrate this, suppose there is a list of 100 schools from which a systematic sample of 7 schools is required. The researcher may select a school at random from the first 14 schools of the list because $7 \times 14 = 98$ is just short of 100. Let the school at serial number 6 be selected from the first 14 schools and then the researcher shall select every 14th school from the list to get a systematic sample of the school at serial numbers 6, 20, 34, 48, 62, 76 and 90.

This method gives a sample which is almost as good as a simple random sample and comparatively much more convenient to draw. However, the researcher must be alert for two kinds of departures from randomness in the arrangement of the population units. If either of the following situations exists, a systematic sample should be avoided.

1. *A trend.* Suppose that a population of 500 teachers is listed according to increasing seniority and a researcher wants to select a sample of 50. First he divides 500 by 50 to get 10 as the size of the interval. Suppose he picks 5 at random from 1 to 10 as a starting number and selects each tenth name at 5, 15, 25, 35, . . . until he draws the desired 50 names. If he picks 10 as starting point, another sample would consist of numbers 10, 20, 30, 40, . . . In the latter possible sample each teacher has more seniority than his counterpart in the former. The means of these two samples would be significantly divergent as regards seniority and other variables strongly associated with it. Similarly many other samples are possible by taking different starting points and there will be a great deal of variability among them. Moreover, the results on any variable strongly related to seniority would be dependent on which of the possible samples happened to be chosen.

2. *A cyclical fluctuation.* Suppose a researcher wants to get a sample of university students utilizing the university library each day. A biased sample will probably be obtained if every seventh day is chosen as a sample unit, for fewer students probably study in the library on Sundays than on week days.

The characteristics of 'trend' and 'cyclical fluctuation' of the listed population unduly increase the variability of the samples and calculations made from such samples will not show the source of variability. The chief advantage of the systematic sample is its simplicity. This method is often far less expensive than unrestricted or simple random simpling method.

Cluster and Multi-stage Sampling

The cluster sampling is a variation of simple random sampling. It is used when the population under study is infinite, where a list of units of the population does not exist, when the geographic distribution of units is scattered, or when sampling of individual units is not convenient for several administrative reasons. School populations which are not completely listed, such as all the 8th class children in a state, may be sampled in groups or clusters. In the development of norms of an intelligence test, for example, it is not feasible for a researcher to select a simple random sample of children in a state, but it is possible to sample schools or groups of such children. Methods of sampling which entail the division of elements of a population into groups, so that the units sampled contain more than one individual of the population, is termed as cluster sampling or area sampling. This method of sampling is economical when the cost of measuring an individual is relatively small and cost of researching him relatively large. In a sample of primary school children, instead of listing all primary school children in a city A, a researcher lists all primary schools in the city, selects at random 20 per cent of these clusters of units, and either uses all the children in the selected schools as the sample or randomly selects children within these schools. In this case the cluster sample approach is combined with random sampling so that there are stages in the sampling process. Such a sampling is called multi-stage sampling.

Multi-stage sampling is used in large scale surveys for a more comprehensive investigation. The researcher may have to use two, three or four stage sampling. For example, in survey type studies, generally mailed questionnaires are used to gather information from people in widely scattered areas and at a relatively low cost. The missing data due to incomplete or partial returns of questionnaires may introduce a bias and a representative sample of data cannot be obtained. In such circumstances, two stage or double sampling is sometimes used to obtain a more representative sample. The researcher may draw a second sample at random from the non-respondents and contact them personally to obtain the desired information with the help of an interview. This method also helps the researcher to check on the consistency of the information obtained from the first sample. Similarly suppose for the purpose of national survey, the research organization like National Council of Educational Research and Training wants to select a sample from all secondary school teachers in India. A simple random sample would be impracticable and so from the states a sample of 5 states could be randomly selected from northern, eastern, southern, western and central regions. From the 5 states all districts could be listed and a random sample of 50 districts selected. From the 50 districts all secondary schools could be listed and a random sample of 100 schools selected. It would not be difficult to campile a list of all school teachers working in 100 schools and a random sample of 700 teachers selected. The successive random sampling of states, districts, schools, and finally of teachers constitute a multi-stage (four-stage) sample.

Multi-stage sampling is comparatively convenient, less time-consuming and less expensive method of sampling. However, an element of sample bias gets introduced because of the unequal size of some of the selected sub-samples. This method is recommended only when it would be impracticable to draw a simple random sample.

Probability Proportion to Size Sample

When the units vary in size, such as schools in respect of total enrolment, it is generally better to select such a sample that the probability of selection of a unit is proportional to its size. Therefore, a school with an enrolment of 800 will get double the chance of being included in the sample compared to a school with an enrolment of 400. Such a sample is known as probability proportion to size or PPS sample.

Errors in Sampling

Even if a sample is properly selected there will be some difference between the estimates obtained from the sample and the actual value in the population. The measures estimated from the samples are called statistics, and will tend to differ more or less from sample to sample drawn from the same population due to fluctuations of sampling. Measures descriptive of a population, on the other hand, are called parameters and are to be thought of as fixed reference value. It is not possible to compute parameters but there is a statistical procedure, to be explained later, to forecast parameters from sample statistics provided certain conditions have been satisfied. The difference between the sample estimate (statistics) and the population value (parameter) is called error.

Sample studies are subject to sampling and non-sampling errors which are of a random and/or of a constant nature. Mouly (1963, pp. 169-70) has classified these errors in a four-way classification table which is shown in Figure 5.1.

	Random	Systematic
Sampling	A	B
Measurement	C	D

FIG. 5.1. Errors of Sampling.

The cells A and B refer to sampling errors while C and D refer to non-sampling ones.

The errors which are due to sampling and of which the average magnitude can be determined are called sampling errors Cell A refers to the unavoidable errors that occur whenever sampling is done and some shift in the sample statistic is caused by selecting at random any individual who may be high, low or average in the trait in question. For example, if the researcher has decided to select a sample of 50 students in a study of achievement motivation and has already chosen the first 49, the 50th student, selected at random, high in achievement motivation, will, therefore, cause some shift in the sample statistics in the upward direction.

Cell B refers to constant errors of sampling. These are also called errors of bias in sampling, that is, errors which do not cancel out but which lean systematically in one or the other direction of the population value. For example, the researcher would probably introduce a bias in his data by drawing a sample from some favoured locations. When the magnitude of systematic errors is large, the data are of limited use for drawing out of the generalisations about the population.

Whatever the procedure of sampling, there are also biases due to faculty tools of measurement, incomplete coverage of sample or non-response, defects in data collection, etc. These biases produce errors which are known as non-sampling errors. These are also called as measurement errors and are represented in Figure 5.1 by the cells C and D. The errors in the cell C refer to random errors of measurement and are due to the unreliability of the testing. On any measuring instrument, most individuals are likely to be mismeasured to some degree due to errors in the procedures of observation, interviewing, coding, non-response, etc. These errors are called as variable response errors and they tend to cancel each other in the long run.

Cell C refers to systematic errors of measurement. It concerns biases which include all the discrepancies between our observations and the quantities we aim to measure of the systematic non-cancelling types. For example, if in the testing of a sample of students for general mental ability, the researcher unintentionally allows an extra time of 2 minutes for the test, there will probably be a systematic tendency for the sample statistic to be higher than it should be and, therefore, there will be a bias in the measurement of mental ability. The size of such biases is unknown in practice. The square of the total error is the sum of squares of sampling and non-sampling errors. That is:

$$(\text{Total error})^2 = (\text{Sampling errors})^2 + (\text{non-sampling errors})^2.$$

This relationship may be illustrated by means of the three sides of a right angled triangle shown in Figure 5.2.

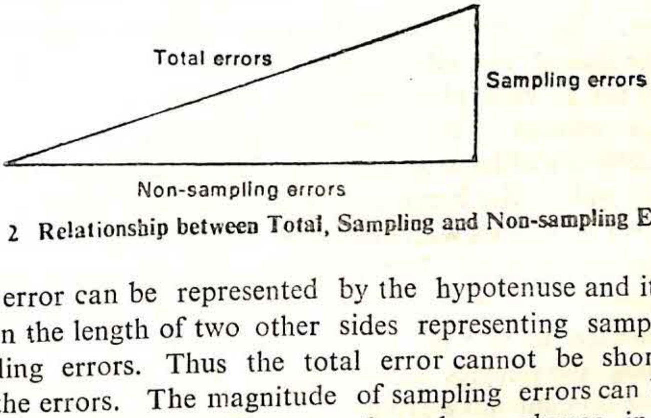


FIG 5 2 **Relationship between Total, Sampling and Non-sampling Errors**

The total error can be represented by the hypotenuse and its length depends on the length of two other sides representing sampling and non-sampling errors. Thus the total error cannot be shorter than either of the errors. The magnitude of sampling errors can be minimised to any fractional value sometimes by a change in sample design, and always by selecting more sampling units, either in the form of more clusters or more individuals. By taking large sample, random sampling errors cancel each other to the point that the sample statistic will tend to stabilize close to the population parameter. Of the non-sampling errors, the random or variable, response errors may be reduced either by taking more observations per individual(s), interviewers or by improving the precision of the methods of observation.

The magnitude of non-sampling errors may be due mostly to bias or systematic error of measurement, which can be minimized only through better survey methods, i.e. through improving the questionnaire or, the field work, or the coding and processing of data. etc. An important category of non-sampling errors in education is composed of the errors of non-response. These arise due to failure in locating some of the individuals of the population selected for the sample or due to their refusal to respond. Non-response errors are also due to the respondents not possessing correct information or due to their giving deliberately biased responses.

Non-response can lead to serious bias in the results. In the surveys conducted by mailed questionnaire the non-response is generally very high. If the researcher has selected a sample of 300 individuals and 90 of them have not responded, then the remedy does not lie in selecting another 90 individuals from the population. Usually the information gathered from respondents differs from that of the non-respondents. The solution is to select a sub-sample of 30 or 40 individuals from the group of non-respondents and to collect information from them using greater efforts and even incurring more expenditure. The wording and the structure of the questions in the questionnaire are also planned with the aim of reducing the incidence of refusals.

In mailed questionnaires the researcher may send reminders to non-respondents after every week or so. With each reminder he gets some responses. There may be a group of persons who would not respond even after several reminders. In such cases one can graphically extrapolate the average observation for this hard core of persons and obtain an unbiased average for the whole group.

The technique of inter-penetrating sub-samples is also used to check the errors which arise because of differences between the interviewers in those survey studies where different interviewers are employed for the collection of data. For example, suppose a state consists of 8 districts and there are 50 units in each district from whom data are to be collected employing 8 researchers. The procedure of inter-penetrating sub-samples would require that all the 400 units are divided at random into 8 groups of 50 each. The groups are assigned to 8 researchers and they would collect data working independently of each other. The data so collected is compared among the 8 researchers and if there is any bias in the data of any of these researchers, the bias would be identified by comparing the 8 averages to make necessary corrections. This method also helps in keeping a psychological check on the researchers.

Choice of Sampling Method

The choice of sampling method depends upon many considerations unique to each individual project. These include the definition of the population, available information about the structure of the population, the parameters to be estimated, the objectives of the analysis including the degree of precision required, and the financial and other resources available for the project.

DATA-GATHERING TOOLS

A researcher will require many data-gathering tools or techniques which may vary in their complexity, design, administration and interpretation. Each tool is appropriate for the collection of certain type of evidence or information. The researcher has to select from the aavilable tools which will provide data he seeks for testing hypotheses. In some situations he may find that existing research tools do not suit his purpose and so he may modify them or construct his own. For this the researcher should familiarize himself with the nature, merits and limitations of the existing research tools; and should also develop skill in the construction and use of each of these research tools.

The major data-gathering tools of research may be classified broadly into the following categories:

1. Psychological Tests.
2. Inquiry Forms.
3. Observation.
4. Interview
5. Sociometric Techniques.

PSYCHOLOGICAL TESTS

Psychological tests are among the most useful tools of educational research. They have been devised to evaluate or measure behaviour in a standardized way for the purpose of selection, classification, prediction and guidance as well as for the evaluation of educational programmes. Tests consists of a series of tasks which the subject is required to perform. They are designed to measure general mental ability or intelligence; special abilities or aptitudes; creativity; achievement; personality traits and adjustment, interests and values.

Cronbach (1964, p. 21.) defines a test as a systematic procedure for comparing the behaviour of two or more persons at a particular time; or one or more persons at a different time. Psychological tests yield objective and standardized descriptions of behaviour which can be quantified by numerical scores. For purposes of educational research, standardized tests are the most commonly used ones because they are considered more objective than the non-standardized ones. A standardized test is one that has specific directions for administration and scoring, a fixed set of test items, and has been administered to representative samples taken from the population for whom the test is intended for the purpose of establishing norms. Norms are usually the average scores made by the representative groups of individuals at different age or grade levels. They make it possible to compare an individual's performance with other individuals belonging to his age or grade level. Thus the chief value of standardized test for research purposes in education lies in their use as tools of comparison. It may be noted that standardized tests are as objective as possible, their scoring is unambiguous and simple.

Selection of Tests

In selecting tests for collecting data in research situations a researcher must evaluate their validity, reliability and usability. These evaluative criteria are considered desirable for a good test.

Validity. The test, as a data collection tool, must produce information that is not only relevant but free from systematic errors; that is, it must produce valid information. In general, a test is valid if it measures what it claims to measure. A test, however, does not possess universal and eternal validity. It may be valid for use in one situation but invalid if used in another. Cronbach (1964, p. 103) says that a test which helps in making one decision in a particular research situation may have no value at all for another. This means that a researcher should not ask the general question. "Is this a valid test ?" The pertinent question to ask is "How valid is this test for the decision I wish to make?" Or more generally, "For what decision is this test valid?"

There are different types of validity including: (1) content validity, (2) criterion-related validity, and (3) construct validity.

Content validity. Content validity is essentially applied only to tests of proficiency and of educational achievement. It is estimated by evaluating the relevance of the test items, in relation to instructional objectives and actual subject matter studied, individually and as a whole. Freeman (1965, p. 91) says that each test item should be a sampling of knowledge or performance which the test purports to measure. Taken collectively, the items should constitute a representative sample of the variable to be tested.

This form of validity is based upon judgement of several subject experts and test specialists, careful analyses of instructional objectives, and the actual subject matter studied. This analysis is rational as well as judgemental and, therefore, the content validity is sometimes also named as rational or logical validity.

The test content is drawn from what has been taught or what is proposed to be taught during an instructional programme. It acts as a blue print from which the test maker prepares his test items. Freeman (1965, p. 91) refers to the type of statistical analyses in the validiting process for the purpose of: (1) determining which items discriminate best between individuals at the upper and the lower levels of performance; (2) determining the per cent that answers each item correctly; (3) determining significance of increases in average scores from one school grade to the next: (4) determining for each items, or for each division (sub test) its correlation with educational progress and with general educational performance (school marks). Thus, according to Kelley *et. al.* (1953), content validation rests first upon an expert analysis of the materials to be sampled (the variable) and the second upon the use of available statistical procedures to refine the original selection of items.

Criterion related validity. In some situations of decision making especially in selection or classification, the decision is based on an individual's expected future performance as predicted from the test score. A test which predicts the kind of behaviour it was intended to predict (e.g. success in a job) is said to possess *predictive validity.* This validity, therefore, refers to the association between present results as indicated by a test and future behaviour; and in order to determine the predictive validity of a test the results from it must be compared with the actual performance in future. If a test is designed to select students for some engineering course, for example, scores on the test must indicate a significant positive correlation with ultimate success in this field. While establishing the predictive validity of a test, a follow up study is required. The researcher administers a test, makes his prediction, tries the treatment suggested by these predictions, and obtains a record of the outcome. The outcome may be the employee's production on a job, the school marks, the doctor's satisfaction with his job after joining the medical profession and the like. A researcher studies predictive validity when his primary interest is in some outcome which he wants to improve by some professional decisions.

In many situations for which tests are developed some cumbersome technique or method of collecting information is already in use. If the existing technique is considered useful for decision making the first question in validation is whether the new test agrees with the present source of information. If they disagree, the new test in spite of its good qualities can certainly not be a substitute for the original technique. The agreement between the test and the existing cumbersome technique for which the test is developed is estimated by an empirical comparison. Both the test and the original technique are applied to the same subjects, and the results are compared. For example, tests developed for some clinical diagnosis are compared with the opinions framed by a psychiatrist. This type of empirical check on agreement is called concurrent validation, because the two sources of information lead to nearly the same results. The validity of the newly developed test thus established is called concurrent validity.

In case of predictive validity, the record of the outcome may be termed as criterion. While investigating concurrent validity the test is proposed as a substitute for some other existing cumbersome technique. The information obtained through this technique acts as criterion. In both the cases the information obtained through the newly developed tests is related to a criterion and, therefore, the two types of validities are also termed as criterion-related validity. It is worth noting that whilst predictive validity refers to the association between present results and future behaviour, the concurrent validity is concerned with the test's ability to provide an estimate of present behaviour.

Construct validity. Construct validity is concerned with the meaning and interpretation of the test scores obtained in terms of psychological or theoretical constructs. A construct is a trait of ability, temperament, or attitude which is hypothesized to explain certain aspects of behaviour such as 'Achievement Motivation', 'Intelligence', 'Creative Thinking' or 'Test Anxiety'. Construct validity thus is concerned not only with the test itself but also with the theory which seeks to explain, or to account for the results which are obtained when the test is used.

Construct validity is established through a long-continued experimentation based on imagination, reasoning and observation. First imagination suggests that a construct X accounts for the test performance. The researcher then reasons "If that is so, then individuals with a high score should have characteristics or attribute Y." An experiment is performed and observations are made, and if the expectation is confirmed, the interpretation is supported. But as various deductions are tested, some of them prove to be inaccurate. The proposed interpretation is then altered either by invoking a different concept, by introducing an additional concept, or by altering the theory of the concept itself. The process of construct validation is the same as that by which scientific theories are developed.

Several procedures are used in the process of construct validation.

1. *Method of factor analysis.* A test which is developed to assess and then to explain the construct or the nature of a trait is correlated to certain other tests. These other tests may be already accepted measures of the triat in question. The patterning of these correlations make it possible to see which tests are measuring some common factors. An examination of the tests clustering together in a single factor may serve to clarify the nature and meaning of the factor.

2. *Method of group differences.* The nature, theory or construct of a trait sometimes suggest that certain kinds of groups should score significantly high or low on the trait and so on the test which is a measure of it. Applying a test to these groups, the hypothesis of difference in the scores of the trait is tested.

3. *Method of studying the effect of experimental treatments or conditions on test scores.* A theory or a construct imply that the expression of a trait in terms of a test score will undergo a change as a result of certain experimental conditions or treatments. For this, a researcher can formulate a network of theory, leading to define predictions. These predictions can be tested. In so far as they are borne out, the validity of the test as a measure of the construct is supported; otherwise we are led to doubt the validity of our test, or construct, or both.

Hence, from the earlier discussion we may conclude that the procedure for assessing the content, criterion-related and construct validity of a test involves one or more of the following:

(*i*) Examining closely the content of the test to make judgements about what is being assessed.

(*ii*) Correlating the test scores with some external criterion which may be record of the outcome like supervisors' ratings, teachers' estimates, examination results or performance on a technique which measures the same factor as the test which is being validated.

(*iii*) Using the methods of factor-analysis, or group differences or making predictions about response to experimental treatments or conditions.

The validity coefficient. The validity of a test is usually reported in terms of validity coefficient. It is the coefficient of correlation between the test and the criterion measures which have been used in the procedure of validation. Lovell and Lawson (1973) recommend that a pp. 139-141) has called it *coefficient of stability*. He has called other coefficients as the *coefficients of equivalence*.

Usability

The usability of a test includes objectivity, economy of time and cost, simplicity and ease of administration, scoring and interpretation in using test for collecting data in a research study. A test should yield objective results, that is, the results should be independent of personal judgement of the researcher using the test. The tests that can be administered in a short period of time are likely to gain cooperation of the subjects and to save the time of all those involved in the test administration. The cost involved in the construction or administration of a test should be limited. The simplicity and ease of administration, scoring and interpretation is an important factor in selecting a test, particularly when the expert advice is not easily available. The tests should be interesting and fascinating to the subjects so as to gain their cooperation.

Researchers should consult *The Mental Measurements Yearbooks* (BUROS) to get themselves acquainted with the available psychological tests. It contains critical evaluation of published tests in addition to the information concerning the names of their publishers, their costs, and their uses. For further detailed information and extensive analysis it would be useful to consult manuals of the tests so as to select tests appropriate for a certain range of age, maturity or ability.

Test Development

The development is an independent, extremely advanced and technical aaea of research and is based on well developed theory of psychometrics. The theory and procedures are being continuously subjected to research, criticism and improvement. Developing the theory or test development is one form of research. Adaptation and improvement of the existing tests and construction of new ones also amount to research activity.

General Procedure of Test Construction

When a researcher sets out to construct a test, there are several factors which may determine his operations and line of action. The factors include the area, the age-group and the grade for which the test is to be developed. The approach of the researcher will depend on whether the test is designed for general purposes or some specific purpose. But in spite of all such factors, there are some general principles and procedures which one has to follow while constructing a test.

Planning. The construction of a test much start by a consideration of the limitations under which the test has to be developed. It includes a detailed set of specifications as to the purpose of the test and the time, cost and sources at the disposal of the researcher or test-maker. The nature of the population for which the test is constructed has to be defined. The length of the test, type and nature of the test-items, and method of scoring the test are also some basic considerations which are to be planned in advance.

Preliminary draft. While making the preliminary draft of the test, the researcher or test-maker mu t consult the existing tests in the concerned area. Such tests are helpful in constructing the items for the test. The test-maker may also create some original items of his own to cover the attribute or trait adequately. The preliminary draft must have more than double the items required for the test. For ability and attainment tests it is necessary to compile a large number of items of suitable difficulty. A rough idea of the difficulty of the items can be obtained by trying out a few items on a small group of subjects from the population. The items are then edited, and carefully worded instructions which indicate briefly the nature and purpose of the test, the nature of the task with a few examples must be supplied with the test. The final manuscript of the preliminary draft is then submitted to experts for their opinion and criticism. It is also worthwhile to administer the final manuscript of the preliminary draft to a small group of subjects from the population and check the answers. It is called 'small-group try out' of the test. This procedure may suggest further modification. After the necessary modifications in the light of experts' suggestions and 'small-group try out', the preliminary draft is printed or cyclostyled. For recording the responses of the subjects a separate answer sheet must also be printed or cyclostyled which may be enclosed with the booklet of the preliminary draft.

The tryout. At this stage the preliminary draft is administered to a large random sample of the population for which the test is constructed. The size of the sample for try out is usually taken as 370 because it helps the test-maker to get the indices of difficulty and discrimination quickly for selecting good items for the final test with the help of the Table developed by J.C. Flanagan.

At the try out stage the time limit should be generous. The test may be so timed that nearly 90 per cent individuals in the sample complete the last item. In case of speeded test, the time limit that produces a good scatter of scores without fatigue should be fixed.

The test booklets along with their answer sheets are collected and scored with the help of a scoring key.

Item analysis. The major objectives of item analysis are the improvement of total score reliability or of total score validity, or both, and the achievement of better item sequences and types of score distributions. Item analysis procedures provide for each item of the test of ability two indices one of its difficulty and another of its power to discriminate between the good and the bad performers on the test. Many kind of such indices have been proposed. One begins by arranging the answer sheets form the highest to the lowest obtained score. From the arranged answer sheets the top 27 per cent and the bottom 27 per cent of the answer sheets are separately taken. Next the proportions of the two groups passing a given item are found. Entering Flanagan's Table with the proportion of successes in the two groups, read the biserial r from the intersecting column and row in the body of the table. The Biserial r will be the *discriminative power* or the *validity index* of an item The mean of the two proportions is the *difficulty index* of the item. Garrett (1962, p. 368) suggests that items with validity indices of 0.20 or more and difficulty indices of 0.40 to 0.60 are regarded as satisfactory. Items having zero or negative validity are useless and must be discarded or they must be carefully examined for ambiguities, inaccuracies and other errors. It is worth noting that the items for non-cognitive tests are selected only on the basis of *validity index*. In such type of tests there is no question of the difficulty value of an item as the subject is required to respond to a series of statements or questions in 'yes' or 'no'; 'agree' or 'disagree'; or in a similar way to indicate his feelings or opinions.

Final draft. The selected items are put in the final draft of the test which is administered to as large a sample as is practicable for estimating the parameters of validity, reliability and norms.

Different types of validity and reliability and the methods for estimating them have been discussed earlier. Norms represent a descriptive framework for interpreting the test score of an individual or a group. A raw score, taken by itself, has no meaning. It gets meaning only when compared with some reference group or groups. Different types of norms are commonly derived from these comparisons. When the population is heterogeneous and covers a wide range of academic attainment or age, norms related to those ranges are developed and are denoted as 'grade norms' and 'age norms'. In these types of norms we give meaning to an individual's test score by determining the age or grade group in which he would be just average. In 'percentile norms' each individual's score is transformed into an equivalent percentile rank. This comparison is based on a single group and indicates what per cent of that group the test score surpassed. The fourth type of norms is called as 'standard scores' in which comparison is again based on a single group and makes use of the mean and standard deviation as a basis for comparison.

Classification of Tests

Psychological tests may be classified in many ways. One distinction is made between power tests and timed or speed tests. A power test is a test in which every subject has a chance to attempt every item. This test has no time limit and the subject goes on attemping test items until he is unable to continue successfully. A speed or timed test is often defined as one in which no subject has time to attempt all items. Such tests limit the time in which the subject has to complete certain tasks.

A distinction between the tests may be made in terms of the way in which they are administered with many subjects at the same time and are named as group tests. Such tests are particularly useful where large number of subjects have to be tested at the same time. In such situations they are preferred because of their economy. These tests are usually simple to administer and score and are widely used for research purposes. It may be noted that such tests should be used for the initial classification or categorisation of subjects rather than for obtaining accurate information about the individuals. Moreover, group tests are likely to yield unreliable results when used with young children, semi-literate or illiterate subjects, or with emotionally unstable subject. For such type of subjects individual tests have been developed. Individual tests are also useful in the situations where we want to obtain a precise and detailed assessment of an individual subject. A researcher needs considerable experience and training in administering many individual tests.

A further distinction can be made between paper-pencil tests and performance tests. Paper-pencil tests require the subject to respond to the test items by writing his replies. They pose questions in the form of sentences, or designs, and require the subject to record his answer either by underlying, ticking or circling one of the alternative answers which are provided, or by writing a word, phrase or sentence in a blank space provided for the purpose. On the other hand, in the performance tests problems are presented in a concrete form and the subject is required to respond not by writing but by manipulating some toys, blocks or picture cards. Such tests are usually administered to subjects individually.

Two other classes of tests are non-standardized teacher-made tests and standardized tests. Teachers make use of their own made tests in classroom situations to assess attainment of students in different school subjects. Such tests are designed for specific use and their reliability and validity are not established by careful statistical controls. By contrast, in standardized tests each item and total scores are carefully analysed. Their validity and reliability are established by careful statistical controls. Such tests provide us norms for various age and grade levels. The content, administration, scoring and interpretation of these tests are standardized.

Tests can also be classified in terms of their purposes that is, the types of psychological traits that they describe and claim to measure. By this standard, we may distinguish five major classes of tests, each with many sub-classes. First of all, are the tests of general mental ability or intelligence; second, the tests of special abilities or aptitudes; third, the tests of creativity; fourth, the tests of attainment; and fifth, the personality measures.

Tests of general mental ability or intelligence. Tests of general mental ability or intelligence measure general ability which enters into performance of all activities and which differs in magnitude from individual to individual. The items in such tests assess the subject's ability to perceive relationships, solve problems and apply knowledge in a variety of ways. Tests of intelligence should not be considered as measure of pure intelligence because the performance on such tests is partly determined by one's background and schooling. The items in an intelligence test are numerous and varied. They test different abilities which are supposed to constitute general intelligence.

Intelligence tests are classified as verbal and non-verbal tests; paper-pencil and performance tests; speed and power tests; individual and group tests.

With all their limitations, intelligence tests are indispensable tools in educational research. They are used for classifying and categorising subjects into various groups. They help a researcher to make decisions about the placement of individuals for school and work, and, therefore help him in making certain predictions. The researcher should be familiar with different intelligence tests. He may come across some standardized tests of intelligence which may serve the purpose of his study. In case no test is available for use with a specific group, the researcher should develop a new test.

A List of Some Intelligence Tests Developed in India and Indian Adaptation of Some Foreign Intelligence Tests

1. *Verbal group test of intelligence* (B.P.T. 6). It is a verbal test of intelligence in Hindi. This test is meant for the children of age 10 plus.

It is a group test and contains 50 items. It can be had from Manovigyan Shala, Allahabad, U.P. The time-limit for the test is 30 minutes.

2. *Verbal group test of intelligence* (B.P.T. 12). It is group verbal test of intelligence in Hindi meant for children of 12 plus. It contains 100 items, and the time required to administer and complete the whole test is 45 minutes. The split-half reliability of the test is 0.97. The test can be had from Manovigyan Shala, Allahabad, U.P.

3. *Verbal group test of intelligence* (B.P.T. 7). This test is a group verbal intelligence test in Hindi meant for children of 13 plus. The split-half reliability of the test is 0.97. Norms are also available in the manual. The test contains 100 items and is completed within 100 minutes. The test is available from Manovigyan Shala, Allahabad, U.P.

4. *Verbal group test of intelligence* (B.P.T. 13). It is a group verbal test of intelligence in Hindi having 100 items and can be completed within 45 minutes. It is meant for children of class VIII of U.P. The test is available from Manovigyan Shala, Allahabad, U.P.

5. *Group test of intelligence in Telugu.* This test has been developed by the Bureau of Educational and Vocational Guidance, Hyderabad (A.P.). It is a verbal group test of intelligence meant for 10 to 16 years of Telugu speaking students. There are 6 sub-tests in the test and the time-limit is different for sub-tests.

6. *P.R.W. tests 1, 2, 4, 8, 20 and 27.* These are group tests of intelligence in English developed by Psychological Research Wing, Ministry of Defence, New Delhi. The reliability coefficients of the tests are above, 0.75.

7. *Pathak vikash.* It is a verbal group test of intelligence in Marathi. The test is meant for the children of age range 9-13 years. It has 8 sub-tests with 100 items in all and can be had from A.K. Pathak, c/o S.T. College, Bombay-I. The reliability and validity coefficients are 0.89 and 0.74 respectively.

8. *Orissa group test of intelligence.* The test has been developed by R.C. Das and G.N. Das. It is a group verbal test of intelligence in Oriya and is meant for the age group of 12 years to adulthood. It contains 70 items and can be completed within one hour. The correlation of the test with different school subjects range from 0.40 to 0.65.

9. *Intelligence tests for high school students in Kannada for Karnatka region.* It is a group verbal test of intelligence in Kannada meant for the age-group of 14 to 18 years. The test contains 14 items to be completed in 60 minutes. The reliability and validity coefficients have

come out to be 0.79 and 0.58 respectively. The test is available from S.T. College, Belgaum,

10. *Maratha buddiwapan patrika "Ma" and "Ga" forms.* It is a group verbal test of intelligence in Marathi meant for high school students whose age ranges from 10 to 17 years. The test has 16 sub-tests in "Ma" form and 12 sub-tests in "Ga" form. The time limit for both the forms of the test is 60 minutes. The test is available from the Education Research Association, Shahpur, Belgaum.

11. *Samooki ka buddi parikshan.* It is a verbal group test of intelligence in Hindi, constructed and standardized by Prayag Mehta. The test is meant for children of age range 12 to 14 years. It contains 11 sub-tests having 60 items in all to be completed in 18 minutes. The reliability and validity coefficients of the test have come to be 0.93 and 0.44 respectively. The test is available from Manasyan, 32 Faiz Bazar, Delhi-6.

12. *Sadharan manasika yogyata pariksha.* It is a verbal group test of intelligence in Punjabi meant for children of 11 to 16 years of age. There are 100 items in the test to be completed within 20 minutes. The reliability and validity coefficients are 0.93 and 0.50 respectively. The test is available from Manasyan, 32 Faiz Bazar, Delhi-6.

13. *Sadharan manasika yogyata pariksha.* It is a group verbal test of intelligence in Hindi developed by S. Jalota. The test is meant for the children of age-range 12-16 years. It has 100 items to be completed within 20 minutes. The reliability and validity coefficients are 0.94 and 0.78 respectively. The test is available from Manasyan, 32 Faiz Bazar, Delhi-6.

14. *Adaptation of Goodard's form board test for measuring intelligence of Gujarat children.* The adaptation of this test has been done by B.C. Patel in Gujarati. It is an individual performance test of intelligence meant for the children of age 6 to 10 years.

15. *Battery of performance tests of intelligence.* It is an individual performance test battery of intelligence developed and standardized by C.M. Bhatia. The battery has 5 sub-tests with different number of items. The time for each sub-test is also different. The reliability and validity coefficients are 0.84 and 0.71 respectively. The test is available from Purohit and Purohit, Shukrawar-Petk, Poona-2. It is meant for the age range of 11 to 16 years.

16. *Adaptation of Raven's progressive matrices test.* The adaptation has been done by Manovigyan Shala, Allahabad, U.P. It is a non-verbal test of intelligence meant for the students of VIII, IX and X classes. The time limit for the test is 20 minutes. The test consists of 60 well-drawn matrices divided into five sets (A, B, C, D, and E) of 12

problems each.

17. *Adaptation of pidgeon non-verbal test.* The test has been adapted by Manovigyan Shalla, Allahabad, U.P. It has been found suitable for the students of class VIII of the age-group 10 to 15 years in India. Time limit for each sub-test ranges from 5 to 6½ minutes.

18. *Hindi adaptation of standford Binet test of intelligence (Scale-L).* It is a mixed type of test (verbal, non-verbal etc). The test is meant for the individuals of age range 2 years to superior adults. Some subtests of the test are timed and some are untimed. It is available from Manovigyan Shala, Allahabad, U.P.

Tests of Special Abilities or Aptitudes

Though intelligence tests seek to measure those mental abilities which are valuable in almost any type of thinking and learning, efficient vocational guidance or placement call for tests specially directed at the more specialized abilities which are of value only in a limited range of tasks. Such type of tests are called as tests of special abilities or aptitudes. According to Freeman (1965, p. 54) an aptitude test is designed to measure a person's potential ability in an activity of specialized kind and within a restricted range. Intelligence tests are often referred to as scholastic aptitude tests as they measure the general mental ability to deal with ideas and symbols without reference to school performance and which enters into the performance of every activity, and, therefore, predict the degree of achievement that may be expected from individuals in various fields.

Aptitude tests are used to predict success in some occupation, academic or training course. These tests have proved of great value for research in educational and vocational guidance. They can be used for predicting rate of learning in special fields of human activity and can avoid considerable wastage and stagnation.

It may be noted that the principles underlying the construction and standardization of intelligence and aptitude tests are same in respect to sampling of performance, population samples, and standardization techniques.

A List of Some Aptitude Tests Developed in India and Indian Adaptation of Some Foreign Aptitude Tests

1. *Clerical aptitude test battery.* This battery has been constructed and developed by the Bureau of Educational and Vocational Guidance, Patna (Bihar). The test is both in Hindi as well as in English. It is meant for pupils of 14 years and above.

2. *A battery of mechanical aptitude tests.* The battery includes three

tests (*i*) Tweezer Dexterity Test; (*ii*) Detriot's Manual Ability Tests; and (*iii*) The Steadiness Tester. It is in Hindi and there is no time limit. The battery is meant for children of age 11 plus to 16 plus. All the tests in the battery are individual and performance tests.

3. *Adaptation of differential aptitude test battery (D.A.T.).* This battery includes tests of (*i*) Verbal Reasoning, (*ii*) Numerical Ability, (*iii*) Abstract Reasoning, (*iv*) Clerical Speed and Accuracy, (*v*) Mechanical Reasoning, (*vi*) Space Relations and (*vii*) Spelling and Grammer. It is available from Manasayan, 32 Faiz Bazar, Delhi-6. The tests included in the battery are highly reliable and valid.

4. *A battery of mechanical aptitude tests.* The battery of the tests is meant for grades VIII through XII. It includes tests of (*i*) knowledge of tools, (*ii*) mechanical comprehension, (*iii*) space relations, (*iv*) form relations and (*v*) mechanical adaptability. The split half reliability coefficients for the tests included in the battery range between 0.90 and 0.98.

The battery has been constructed by A. Sharma and is available from Central Bureau of Educational and Vocational Guidance, Department of Psychological Foundations, N.I.E; 33, Chhatra Marg, Delhi.

5. *Aptitude tests for secondary school teachers.* The test has been constructed by M.M. Shah. It is a verbal group test to measure the aptitude of prospective secondary school teachers. The time limit for the whole test is about 100 minutes. The split-half reliability of the test is 0.88. The test is available from Faculty of Education and Psychology, M.S. University, Baroda.

6. *Adaptation of Bennett's mechanical comprehension test.* The test has been adapted by the Institute of Vocational Guidance, Government of Maharashtra, Bombay. It is a group verbal test. The test contains 60 items and is meant for children of class X and onwards. The test has been validated with the marks in the annual examination for civil engineering degree (N=36, r=0.34); with marks for mechanical and electrical engineering degree (N=45, r=0.54) and with marks for diploma course in engineering (N=70, r=0.42).

Tests of Creativity

Guilford (1967) have brought to the fore the existence of two distinct types of thinking abilities, namely convergent thinking and divergent thinking. Convergent thinking involves the generation of ideas and facts from the known information and in tests of convergent thinking (tests of general intelligence and aptitudes) the subject is asked to define a word, solve an arithmetical problem, find the next number to continue a series, etc. Divergent thinking involves generating of novel responses to situations; responses that are original, unusual and varied. This thinking generally goes by the name of creativity.

In recent years a number of tests to measure creativity have been developed. Guilford (1967) and his associates developed a number of tests which call upon the subject to generate multiple responses to a problem. The responses are scored in terms of fluency (number of responses), flexibility (number of different categories of responses), and originality (number of rare responses given by few other respondents). Torrance (1966) has developed tests oriented more towards young children and towards graphic or pictorial or verbal response. Correlations of different measures of creativity with conventional tests of intelligence have been reported to be low but positive. The extent to which creativity tests are predictive of academic achievement in or out of school is still a matter of controversy and, therefore, Thorndike and Hagen (1970, p. 377) recommend that the tests of creativity must still be thought of as research tools, rather than as instruments that are of proven value for use either in schools or in industry.

A List of Creativity Tests Developed in India and Indian Adaptation of Some Foreign Creativity Tests.

1. *Rachna shakti parikshan.* This test has been developed by N.S. Chauhan and Govind Tiwari. It is in Hindi and measures five factors: creative production, fluency, original power, flexibility and ingenious solution of problems. The test is further divided into sub-factors, units and sub-units. The time limit is liberal for thirteen units and is fixed for the rest of the forty-seven units. The total test takes about three hours for completion. The reliability and validity have been calculated separately for each factor. Test is available from Agra Psychological Research Cell, Tiwari Kothi, Belanganj, Agra.

2. *Passi test of creativity.* The test has been developed by B.K. Passi and is both verbal as well as non-verbal. It is both in Hindi and English. The test contains six sub-tests: seeing problem, unusual uses, consequences, inquisitiveness, square puzzle and block test of creativity. The test is available from National Psychological Corporation, 4/2?0, Bhargava Building, Kacheri Ghat, Agra.

3. *Verbal test of creative thinking.* The test has been developed by Baqer Mehdi in Hindi as well as in English. It is meant for all grades of students except primary. It has been standardised on large samples. Samples included both rural and urban population. It provides three factor scores: fluency, flexibility and originality. The total time for

the completion of the test in 48 minutes. The test is available from National Psychological Corporation, 4/230, Bhargava Building, Kacheri Ghat, Agra.

4. *Non-verbal test of creative ihinking.* The test has been developed by Baqer Mehdi and is available in Hindi and English. It provides two factor scores only: elaboration and originality. The test is also available from National Psychological Corporation, 4/230, Bhargava Building, Kacheri Ghat, Agra.

Tests of Attainment or Achievement

Tests of attainment or achievement have been constructed to measure present performance in relation to a skill or knowledge which has been acquired as a result of training. They are designed to measure an individual's quality of learning in a particular school subject or course of study at the end of the instruction. Achievement tests are not designed to provide an indication of future success, but Beggs and Lewis (1975) have indicated that such tests are found to be as good for predicting success as aptitude tests because of the considerable overlap in the content of the two.

Achievement tests are developed keeping in view their use and interpretation. If the test is to assess whether a student possesses required skills at a particular time and whether those skills can be identified, then an achievement test in terms of a given criterion should be developed. If, on the other hand, the test is to assess the status of student's relative position to some identified group, then an achievement test in terms of some norm should be developed. The former type of the achievement test is designated as mastery or criterion-referenced test and the latter as discriminatory or norm-referenced test. The purpose of criterion-referenced test is to distinguish the student who have and students who have not mastered the minimal essentials; while norm-referenced test is designed to differentiate, as much as possible, the students who have reached different levels of achievement. The teaching model in the criterion-referenced measurement is based on the concept that students are different primarily in the speed with which they can learn rather than in the amount they can learn. Thus the amount that is to be learned in a classroom is fixed for all the learners, but the rate of learning is allowed to vary. The norm referenced measurement compares a student's performance to the class or school average rather than to some predetermined criterion. It refers to the relative achievement of other students to interpret performance. The norm referenced measurement advocates a different teaching model which allows the amount of material learned to vary greatly from student to student, but expects all students to proceed through the subject matter at about the same pace. Here no criterion level for a minimum performance is established in advance. In an influential work on testing measurement in the classroom, Scannell and Tracy (1975) have urged to specify a criterion level of minimum acceptable performance for each objective in a criterion referenced model, but with the norm referenced model no criterion is necessary.

Achievement tests may be classified as traditional or essay-type and new-type or objective tests. New type or objective tests include multiple-choice, true-false, completion, matching and short-answer items.

Achievement tests require only content validity. The reliability of such tests may be found out either by using test-retest method or split-half method. Equivalent forms reliability seems to be of little value with respect to achievement tests because it is concerned with two different forms of a test giving consistent results.

Personality Measures

Personality measures are mainly classified into projective and non-projective techniques.

Projective techniques are used to assess areas of personality that cannot be measured easily by other means, or areas in which direct questions are not likely to be answered truthfully. These techniques disguise their purpose in such a way that the individual unconsciously projects his personality through his responses to a series of fluid and weakly structured stimuli. The subject is presented with ambiguous stimuli such as pictures, ink blots, words and unfinished sentences; and is asked to describe or to interpret what he sees or to complete sentences.

There are few projective media, materials and procedures which have been standardized and are widely used. A scoring procedure has also been devised for these techniques. But the interpretation of the responses often demand a high degree of skill and a prolonged period of training on the part of the person using them. The *Rorschach Inkblot Test, the Thematic Apperception Test* (*T.A.T*), *Children Apperception Test* (*C.A.T.*) *and Rosenzweig's Picture Frustration Test* are some projective techniques which are extensively used in personality assessment.

There are some projective tests which have been adapted in India. Manovigyan Shala, U.P. Allahabad has adapted Rorschah Ink Blot Test meant for individuals of all ages; Bellak's Children Apperception

Test meant for children upto 10 years; Murray's Thematic Apperception Test meant for individuals of 10 years and above, and word Association Test meant for individuals of all ages. Psychological research wing of Ministry of Defence, New Delhi has also adapted Thematic Apperception Test II; Thematic Apperception Test III; Word Association Test V and Word Association Test VI meant for adults. There is also an adaptation of Bellak's Children's Apperception Test by Uma Choudhry. It has been published by Manasayan, 32 Faiz Bazar, Delhi.

The non-projective techniques of personality measurement include personality inventories. Freeman (1965, p. 556) has classified personality inventories into five types: those that (1) assess specified traits (for example, ascendance, sociability, self-confidence, perseverance); (2) evaluate adjustment to several aspects of the environment (home, society, school); (3) classify into clinical groups (paranoiac, psychopathic personality); (4) screen persons into two or three groups (psychosomatic disorders versus normal); (5) evaluate interests, values and attitudes. (vocational interests, scientific and economic values, attitude towards religion).

An inventory is constructed in the form of a questionnaire. It consists of a series of questions or statements to which subjects respond by answering 'yes' or 'no'; 'agree' or 'disagree'; or in some similar way indicate his opinions or feelings. The subject is required to indicate preferences or to mark those items that describe his typical behaviour. Valid results cannot be obtained with the help of inventories if the subject does not provide truthful and frank replies.

The inventories have been mostly used for the investigation of interests and values, but increasingly their use is being made in the assessment of personality traits and adjustment. The predictive validity of their results will be low if the subjects tested are below seventeen years of age, since studies by Strong (1943) suggest that interests are stable until this age is attained.

A list of some Inventories Developed in India and Indian Adaptation of Foreign Inventories

1. *Vyaktitiva parkash prashnavali*. This personality inventory is developed to measure adjustment in different areas: Home, Health, School, Emotional and Social. It is in Hindi and can be used with students of age ranging from 11 years to adulthood. The split half reliability of the inventory has come out to be 0.87. It has been validated against Asthana's abjustment Inventory and the coefficient of validity has come out to be 0.80. The inventory has been developed

by M.S.L. Sexana and is avilable from Central Bureau of Educational and Vocational Guidance, Department of Psychological Foundations, 33-Chhatra Marg, Delhi.

2. *Sixteen personality factor questionnaire.* This inventory has been constructed and standardized by R.B. Cattell. It measures sixteen personality factors, namely, reserved vs. outgoing, less intelligent vs. more intelligent, etc. The inventory is meant for adults and is available from the Psycho-Gentre, T-22, Green Park, New Delhi.

3. *High school personality questionnaire* (*HSPQ*). This inventory measures 14 personality traits. It has also been developed by R.B. Cattell for the children of age range 12-16 years. The inventory is adapted in Hindi by Psycho-Centre, T-22, Green Park, New Delhi.

4. *Edwards personal preference schedule* (*EPPS*). This inventory has been designed to assess the relative strength of 15 manifest needs selected from Murray's need system. It is meant for individuals of age range 13 years to adulthood. It has 225 paired items and has been adapted in Hindi by R.B. Bhatnager, presently working in Meerut University and is available from him.

5. *Maudsly personality inventory* (*MPI*). This inventory has been adapted in Hindi by S.S. Jalota and S.D. Kapoor. It has been designed to measure extraversion-introversion and neuroticism-stability. The inventory is available from Psycho-Centre; T-22, Green Park, New Delhi.

6. *Adjustment inventory.* The inventory has been designed to assess adjustment in different walks of life, in Hindi, by H.S. Asthana. It has been developed for high school and college students. The reliability of the inventory is 0.97. It has 42 items and is available from Psychological Laboratory, Lucknow University, Lucknow.

7. *Vocational preference record.* The inventory is developed for assessing interest in different areas. It has been developed by Manovigyan Shala, Allahabad, U.P. for students of Class X and XIII. The inventory has ten scales and its time limit is 30 minutes.

8. *Chatterji's non-language preference record.* This is a non-verbal inventory meant for Hindi speaking students of class VIII and upwards. It has ten sub-tests. The inventory is available from Indian Statistical Institute, 203 B.T. Road, Calcutta.

9. *Adaptation of kuder preference record.* The adaptation has been done in Hindi by the Bureau of Educational and Vocational Guidance, Patna. It is meant for the assessment of interests of children of age 16 and onwards. The time limit is one hour.

10. *Study of values.* This inventory was originally developed by All-port, Vernon and Lindzey. Its adaptation in Hindi has been done

by R.B. Bhatnagar of Meerut University, Meerut. It measures six values: theoretical, economic, aesthetic, social, political and religious.

11. *The money problems checklist.* It is an inventory which is used by a student in reporting his own problems of adjustment. The subject is asked to indicate on the checklist the problems that trouble him. The inventory has been used as research device to identify and describe the nature of the problems facting individuals and groups of individuals in a school.

INQUIRY FORMS

Inquiry forms are a set of data-gathering research tools which make use of properly designed proformas for inquiring into and securing information about certain phenomena under investigation. Tools included in this category are questionnaire, schedule, checklist, rating scale, score card and opinionnaire or attitude scale.

Questionnaire

A questionnaire is a device consisting of a series of questions dealing with some psychological, social, educational, etc; topic(s) sent or given to an individual or a group of individuals, with the object of obtaining data with regard to some problems under investigation. Goode and Hatt (1952, p. 33) say that in general the word 'questionnaire' refers to a device for securing answers to a series of questions by using a form which the respondent fills in himself. Barr *et al.* (1953, p. 65) define questionnaire as a systematic compilation of questions that are administered to a sample of population from which information is desired.

Questionnaire is a popular means of collecting all kinds of data in research. It is widely used in educational research to obtain information about certain conditions and practices, and to inquire into opinions and attitudes of an individual or a group. A questionnaire is either administered personally to a group of individuals or it is mailed to them to save a great deal of time and money in travel. In the former situation, the person administering the tools has an opportunity to establish rapport with the respondents, to explain the purpose of the study to the respondents, and to explain the meaning of questions to the respondents that may not be clear to them. In the latter situation, the mailed questionnaire is probably both the most used and most abused data-gathering research tool. It is mostly used when the individuals from whom we desire information cannot always be contacted personally without the expenditure of great deal of time and money in travel. The range of administration of a mailed questionnaire may be national or international. A mailed questionnaire is mostly criticized because of an unfavourable reaction which it generates among the respondents to answer questions from strangers on a subject in which respondents are least interested. Best (1977, p 157) calls it "as the lazy man's way of gaining information". He is of the opinion that the unfavourable reaction among the respondents is intensified when the questionnaire is long, the subject trivial in importance, the question items vaguely worded, and the form poorly organized. As a result of this, a small proportion of mailed questionnaires are returned and, therefore, the data obtained are of low validity. Best (1977, p. 167) cautions that a mailed questionnaire may be used with a group of respondents who have a genuine interest in the problem under investigation, who know the sender, or who have some common bond of loyalty to a sponsoring institution or organization.

Classification of Questionnaires

The questionnaires can be classified in terms of the nature of the questions which are used. Questions may be asked in a closed or an open form. The researcher may use one type exclusively or both in combination. Questionnaires that call for short or check responses are known as closed form or restricted type. They include a set of questions to which respondents can reply in a limited number of ways. The respondent is invariably permitted to reply only with 'yes' or 'no', or 'no-opinion', or is requested to select an answer from a short list of possible responses. He is asked to place a tick (√) mark in a space provided on the answer sheet or he may be requested to underline a response. Sometimes he is asked to insert brief answers of his own. For certain type of information the closed type questionnaire is useful because it is easy to respond, takes little time to fill out, is relatively objective and is fairly easy to tabulate and analyse. While using closed-type of question item, it is advisable to provide for unanticipated responses by allowing an 'open' category of response with a request, 'please specify' or 'kindly mention' which enables the researcher to properly tabulate and classify such responses. The following example illustrates such type of question item:

Please tick mark(√) the reasons given below for not introducing the 'Grade system of Marking' by your school.

(*a*) Non-availability of administrative guidance and support ( )
(*b*) Non-availability of academic guidance and support ( )
(*c*) It is time-consuming ( )

(*d*) It is not effective in overcoming the defects of Numerical Marking. ( )

(*e*) It cannot be changed and adopted according to the local/institutional needs. ( )

(*f*) Any other, please mention.

... ...
... ...
... ...
... ...
.. ..

The open-form or unrestricted type of questionnaire calls for a free response in the respondent's own words. The form of the questions is unstructured and no clues are provided to the respondent. The open form of question provides for greater depth of response and the greatest advantage of this type of questions is freedom that is given to the respondent to reveal his opinion and to clarify his responses. However, the responses to such type of questions are sometimes difficult to tabulate, organize and interpret. The following example illustrates this type of question:

State the reasons for not introducing the 'Grade-System of Marking' in your school.

Construction of Questionnaire

Constructing a good questionnaire requires both ability and perseverance on the part of the researcher. He should not use a 'shotgun approach' by attempting to cover his field of investigation broadly in the hope that some of the responses will provide answers for which he is searching blindly. There are some hints which a researcher may consider useful while constructing, administering and analysing a questionnaire.

1. *Purposes of the question naire*.. A good questionnaire must serve two major purposes. First, it must translate the objectives of an investigation into specific questions, the answers to which will provide the data necessary to test the hypotheses and explore the area defined by the objectives. For this each question must convey to the respondent the meaning of each objective so that the responses thus obtained can be analysed and interpreted properly in the light of the research objectives. The statement of the research objectives and the specification of the data required to meet those objectives must precede questionnaire construction. Secondly, the questionnaire must motivate the respondents to communicate the required information. It

is essential to include a courteous and carefully constructed covering letter to explain the purpose of the study. The covering letter should assure the respondent that delicate information will be held in strict confidence. The respondents will feel favourably disposed towards the inquiry if the purpose and nature of the study are stated on the cover letter.

2. *Language*. In the construction of a questionnaire the primary criterion for the choice of language is that the vocab lary and syntax should offer maximum opportunity for complete and accurate communication of ideas between the researcher and respondent. Since the researcher has to depend on written language alone to get the required information, he has to be careful while phrasing the questions. There are some principles that might be employed to make the questions more precise. According to Best (1977, pp. 160-162) following are the principles of making the questions more precise:

(*a*) *Define or qualify terms that could easily be misinterpreted.* Such common words like 'how much', 'new', 'curriculum', 'democracy', 'progressive education', 'cooperation', and 'integration' are all liable to misinterpretation. To the question, "What work are you engaged in now?" the respondent may be tempted to answer, "Filling out your foolish questionnaire."

(*b*) *Be careful in using descriptive adjective and adverbs that have no agreed upon meaning.* This fault is frequently found in rating scales as well as in questionnaire, 'Frequently', 'occasionally', and 'rarely' do not have the same meanings to different persons. One respondent's 'occasionally' may be another's 'rarely'.

(*c*) *Beware of double negative.* For example, the use of double negative in "The State Government should *not* provide aid to the educational institutions in which education is *not* equal, regardless of race, creed, or caste" is misleading.

(*d*) *Be careful of inadequate alternatives.* In the question: "Are you married? Yes—No—" The alternatives Yes/No are inadequate. Does this question refer to present or former marital status? How would the person answer who is widowed, separated, or divorced?

(*e*) *Avoid the double-barreled questions. Divide them into two questions.* Do you believe that gifted children should be placed in separate groups for instructional purposes and assigned to special schools?

In this question one might agree on the advisability of separate groups for instructional purposes, but be very much opposed to the assignment to gifted students of special schools.

(*f*) *Underline a word if you wish to indicate special emphasis.* For example, "Should all schools offer modern foreign language?" will get the desired responses more easily than if the word modern was not emphasized by underlining.

(*g*) *When asking for rating or comparisons a point of reference is necessary*. For example, "How would you rate this student teacher's classroom teaching? Superior—Average—Below Average—" is vague until the respondent is told with whom is the student-teacher to be compared—an experienced teacher, other student teachers, former student teachers—or should the criterion be what a student teacher is expected to be able to do?

(*h*) *Avoid unwarranted assumptions*. For example, to the question, "Are you satisfied with the salary raise that you received last year?" a 'no' answer might mean that the respondent did not get any rise, or that he is not satisfied with the rise that he got.

(*i*) *Phrase questions so that they are appropriate for all respondents*. For example, "How many days per week do you send your student-teachers for practice teaching?" may not suit all respondents as some training colleges have continuous, some block, and some a combination of block and continuous teaching practice.

(*j*) *Design questions that will give a complete response*. The question, "Do you read The Hindustan Times? Yes—No" would not reveal much information about the reading habits of the respondent. It should be followed by an additional item.

(*k*) *Provide for the systematic quantification of responses*. To ask respondents to rank, in order of preference, a specific number of responses from a given list of items permits a convenient way of tabulation by inverse weightings:

Ist Choice	5 points
2nd Choice	4 points
3rd Choice	3 points
4th Choice	2 points
5th Choice	1 point.

The total weighted scores would be the score for the particular item.

(*l*) *Consider the possibility of classifying the responses yourself, rather than having the respondent choose categories*. For example, if a respondent is asked to classify his occupation in the following categories, the result might be quite unsatisfactory.

(*i*) Skilled labour ———
(*ii*) Unskilled labour ———
(*iii*) Clerical work ———
(*iv*) Managerial work ———
(*v*) Profession ———
(*vi*) Proprietorship ———

It is likely that by asking the respondent one or two short questions about his work, it could be classified more accurately.

(*i*) At what place do you work ?

(*ii*) What kind of work do you perform ?

3. *Information level of the respondents.* The assumptions about the expertness of the respondent in a particular field or the amount of information he possesses should not be unrealistic. The information elicited by the questionnaire must tie into the respondent's present level of information.

The questions above the information level of respondents may result in resentment and embarassment among the respondents. This problem is sometimes referred to as 'expert error', that is, the error of ascribing to the respondent a degree of expertness in a particular field which he does not possess.

4. *Social acceptance of responces.* The questions must provide the respondent a range of responses which meets his criteria of social acceptability. A question constitutes a threat to the respondent's ego if he is required to give an answer which he feels is socially unacceptable. The respondent should not be confronted with the necessity of giving a socially unacceptable response to a question. The annoying or embarassing questions must be avoided.

If the information desired from the respondents is of intimate or delicate nature, consider the possibility of providing for anonymous responses to get objective and reliable information. Even if the respondent's name is necessary for purposes of classification, it is essential to assure the respondent that his responses will be kept strictly confidential.

5. *Leading questions.* The questions should be objective with no leading suggestions as to the most appropriate response. For example, in a question "Would you say that you are in favour of co-education in secondary schools?", it is easier for the respondent to answer 'yes' than 'no'. In answering 'yes' he is merely agreeing with the language of the question. It is more difficult to respond, "no", since this response seems to contradict, or at least goes counter to the ideas of the person who worded the question. There are some words which involve respondents emotionally, either favourably or unfavourably, in a particular culture. It is advisable not to make use of such emotionally 'loaded' words while phrasing questions of a questionnaire.

6. *Sequence of questions.* First, the questions should be limited to a single idea or to a single reference. Secondly, the questionnaire maker needs to give thought to the arrangement of the questions in a questionnaire. The questions should be so arranged that they permit

the ideas of the respondents to flow logically. The sequence of questions must facilitate the easy progress of the respondent from item to item and it should lead the respondent to anticipate the next question. Questions should be presented in a good psychological order by adopting the 'funnel approach'. This is a procedure of asking the most general or the most unrestricted questions first and following it with successively more specific and restricted questions. This order helps the respondent to organize his own thinking and motivates him to respond logically and objectively. The following sequence of questions on the 'introduction of the innovation of internal assessment', illustrates the funnel approach:

(*i*) Is the innovation introduced by your school Yes/No

(*ii*) If yes, please mention the year from which it was introduced ....

(*iii*) Is the innovation still continuing in your school? Yes/No

(*iv*) If yes, please tick mark (√) how far the innovation has been useful in overcoming the defects of purely 'External Public Examination'

- (*a*) Greatly useful ( )
- (*b*) Considerably useful ( )
- (*c*) Somewhat useful ( )
- (*d*) Little useful ( )
- (*e*) Not useful ( )
- (*f*) Not known ( )

(*v*) If no, please mention the year from which the innovation was discontinued ———

(*vi*) Please tick mark (√) the reasons given below for discontinuing, or not introducing the innovation:

- (*a*) There is persistent over-assessment of the candidates by the teachers for extra-academic consideration such as personal likes and dislikes, caste and community, payment of money for private tuition or through other forms of subterfuge ( )
- (*b*) The innovation lowers the standard of education ( )
- (*c*) The innovation makes the standard different from school to school ( )
- (*d*) The innovation is time-consuming ( )
- (*e*) The innovation is highly expensive and requires large expenditure ( )
- (*f*) Any other, please mention

....

....

....

(*vii*) On the basis of your experience and scholarship you may wish to provide any additional information about the innovation which is not covered above. If so, please mention.

....

....

....

7. *The form or type of questions.* Another important consideration that weighs in the matter of constructing a questionnaire is that of 'form or type' of questions. As discussed earlier, the questionnaire may contain closed or open type of questions. Each type of these questions has its merits and limitations and the questionnaire framer must decide which type is more likely to supply the information required. Kahn and Cannell (1957) are of the opinion that the closed type of questions are well adapted to situations in which (1) there is only one frame of reference from which the respondent can answer the question; (2) within this single frame of reference, there is a known range of possible responses; and (3) within this range, there are clearly defined choice points which accurately represent the position of each respondent. In the open type of questions, the respondent is encouraged to structure his answer as he wishes and it permits the respondent to state his own frame of reference when this is desirable. In some research situations it is useful to include both the open and the closed type questions in combination.

8. *Length of the questionnaire.* A questionnaire should not be any longer than is necessary. The total number of questions must not be so many as to tire or bore the respondents. If too many questions are asked and the respondent becomes tired, the questions at the end of the series may not be well answered. If it is necessary to include a larger number of questions, then it is advisable to have two separate questionnaires.

9. *Experts' opinion.* It is advisable to get all the help from experts for planning and constructing a questionnaire. Questions should be submitted to the experts for criticism and modified accordingly.

10. *Preliminary tryout of the questionnaire.* No matter how careful the questionnaire maker has been in phrasing his questions and designing his questionnaire, he needs to try them out with a few representative samples of the respondents before launching into the actual investigation. The first purpose of the pre-test is to examine the questionnaire from the research point of view. It is suggested that interviews with respondents should be conducted to see whether the responses fulfil the objectives of the investigation. Sometimes the tryout calls for major revision of the questionnaire and several tryouts

are required until a workable questionnaire is developed. The second purpose of the tryout is to determine the extent to which the questionnaire fulfils the following two criteria:

(*i*) Does the questionnaire promote the congenial and appropriate relationship with respondents?

(*ii*) Do respondents understand the questions without having to be explained or reworded?

11. *Validation of a questionnaire*. The validation of a questionnaire utilizes the same principles and procedures as the validation of any tool of measurement. Each question of the questionnaire must be related obviously to the topic under investigation. In some situations, the questionnaire is validated against the actual behaviour, which acts as the external criterion. This is done by relating question responses with the actual behaviour.

12. *Reliability of a questionnaire*. The test-retest method is the feasible approach to the working out of the reliability of questionnaires. Besides, consistency of the question responses, the accuracy of responses and the comparison of the accuracy of the question responses to the interview may also be taken note of while talking of the reliability of questionnaire.

13. *Administration of a questionnaire*. The questionnaire can call for written or oral responses. Some questionnaires are designed with the intention that they should be answered in writing, others are designed in a way which require an oral response. Questionnaires which require written replies are administered either by post or in a group situation such as a group of students in a class. It is always advisable to send a self-addressed stamped envelope alongwith the questionnaire to exert a moral pressure for return of duly filled questionnaire administered by post. In case of delay, one or two reminders may be helpful. Questionnaires which call for oral replies are usually presented to one subject at a time in an interview.

14. *Analysing and interpreting questionnaire responses*. Quantification of data obtained by the questionnaire is generally achieved through tabulation and counting. The totals are converted into proportions of percentages. Calculation of contingency coefficients of correlation is often made in order to suggest probability of relationship among data. Computation of Chi-square statistics is also advisable.

Limitation of a Questionnaire

1. The questionnaire cannot be used with children and illiterates.

2. The return of the mailed questionnaire is often as low as 40 per cent to 50 per cent. As a result of this poor response, the data obtain-

ed are often of limited validity. The respondents who return the questionnaires may not be representative of the entire group. It will make the sample a biased one and thus vitiate the findings.

3. Sometimes respondents may not like to respond in writing to the questions of intimate and confidential nature or to the questions involving certain controversial issues.

4. It is sometimes difficult to formulate and phrase questions on certain complex, delicate and intricate problems.

5. There is no check on a respondent who misinterprets a question or gives incomplete or indefinite responses.

6. Sometimes the respondent may modify his earlier original responses to the questions when he finds that his responses to latter questions are contradicting the previous ones.

Schedule

Schedule is a device consisting of a set of questions which are asked and filled in by an interviewer in a face to face situation with another person. It differs from the questionnaire in that the former is administered personally to a respondent or a group of repondents while the latter is usually mailed. The schedule has many advantages over the questionnaire. It is administered personally and therefore it provides opportunity to the researcher to establish rapport with the respondents. This helps the researcher to explain the nature and purpose of the investigation and to make the meaning of the questions clear to the respondents if they misinterpret a question or give incomplete or indefinite responses. The schedule also economises time and expense of the investigation. It ensures almost all complete and usable returns.

A schedule has some limitations also. It may not be possible to contact personally all the respondents either individually or in a group.

The procedure of constructing a schedule, the analysis and interpretation of data gathered through it are not very different from those of a questionnaire and need not be repeated here. The administration of a schedule in a face to face situation, collectively or individually, is different from the administration of a questionnaire, but it is so much like the administration of an interview to be discussed later.

Checklist

A checklist is a simple device consisting of a prepared list of items which are thought by the researcher to be relevant to the problem being studied. After each item a space is provided for the observer to indicate the presence or absence of the item by checking 'yes' or 'no', or a type or number of items may be indicated by inserting the appropriate word or number. A checklist draws the attention of the observer to relevant factors and enables him to record the data quickly and systematically.

Thus the responses to the checklist items are a matter of 'fact', not of 'judgment'. The checklist is an important tool in gathering facts for educational surveys, that is for checking of library, laboratory, game facilities, school buildings, textbooks, instructional procedures, etc. Checklists are sometimes used in the form of a questionnaire which are completed by the respondent rather than by the observer.

Construction of a Checklist

The researcher should examine carefully the checklists which have already been prepared and used by other researchers in various fields of educational research. He should then determine the items on which he requires information for his investigation. The items thus determined may be arranged in a logical and psychological order. There are various ways of writing and arranging the items in a checklist.

Kempfer (1960) has suggested four ways and the researcher may make use of all or some of them to serve his purpose best.

1. The form in which the observer or respondent is asked to check all items found in a situation. For example, put a tick mark (√) in the blank provided before each game played in your school :

——— Football
——— Hockey
——— Cricket
——— Volleyball
——— Basketball

2. The form in which questions with a 'yes' or 'no' are asked to be encircled, underlined or checked in response to the items given. For example:

Does your University have a Teachers' Union ? Yes/No

3. The form in which items are positive statements and the respondent or observer is asked to put a tick mark (√) in the space provided on the right of the item.

Put the sentence as For Example:

Our School has a Students' Union. ( )

4. The form where items can best be put in sentences and the

observer or respondent is asked to check, underline or encircle the appropriate word/words.

For example:

The School organizes debates weekly, fortnightly, monthly, annually, irregularly.

The items of the checklist should be phrased in such a way that they are discriminative in quality. It will increase the validity of the checklist. A preliminary tryout of the checklist may also prove helpful in making the tool more objective.

Analysis and Interpretation of Check-list Responses

The tabulation, quantification and interpretation of the checklist responses is done in very much the same way as that of the questionnaire responses.

Rating scale

An observer may be asked to judge the behaviour he observes and classify it into categories. This is essentially the task he performs when completing a schedule, but he can also be asked to give a numerical value or rating to his judgments. By 'rating' is meant the judgment of one person by another. In the words of Barr, Davis and Johnson (1953):

> Rating is a term applied to expression of opinion or judgment regarding some situation, object or character. Opinions are usually expressed on a scale of values.

Rating scale refers to a scale with a set of points which describe varying degrees of the dimension of an attribute being observed.

The rating-scale procedures exceed all psychological-measurement methods, that depend upon human judgment, for popularity, use and ease of administration. They are used in the evaluation of individuals, their reactions and in the psychological evaluation of stimuli. Rating scales are also used to record quantified observations of a social situation. They may be used to describe the behaviour of individuals, the activities of an entire group, the changes in the situation surrounding them, or many other types of data.

Forms of rating scales. A number of rating techniques have been developed which enable the observers to ascribe numerical values or ratings to their judgments of behaviour. According to Guilford (1954, p. 263), these techniques have given rise to five broad categories of rating scales:

1. Numerical scales
2. Graphic scales

3. Standard scales
4. Rating by cumulative points
5. Forced choice ratings.

Numerical scales. In the typical numerical scale, a sequence of defined numbers is supplied to the rater or to the observer. The rater or the observer assigns to each stimulus, to be rated, an appropriate number in line with these definitions or descriptions. One example of such a scale that Guilford (1954, p. 263) used in obtaining ratings of the affective values of colours and odours is as follows:

10 Most pleasant imaginable
9 Most pleasant
8 Extremely pleasant
7 Moderately pleasant
6 Mildly pleasant
5 Indifferent
4 Mildly unpleasant
3 Moderately unpleasant
2 Extremely unpleasant
1 Most unpleasant
0 Most unpleasant imaginable.

In such type of scales sometimes zero is placed at the 'indifferent' category and negative numbers below it. But Guilford (1954, p. 264) does not favour the use of negative rating numbers as this type of scale may be unnatural to those observers or raters who are not well versed in algebra. The use of zero may also tend to suggest a break in the scale and thus destory 'continuity' of the scale.

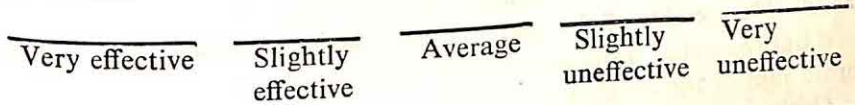
It has been seen that observers or raters usually avoid terminal categories. If such categories (0 and 10) are not included, observers or raters would tend to avoid categories 1 and 9 and thus the range of ratings gets shortened. To avoid this shortcoming it is suggested to expand the scale beyond the categories which a researcher wants to include in his scale. For example, if a researcher wants an effective scale of seven points, he may make use of additional two categories so that desired dispersion of seven point rating is achieved.

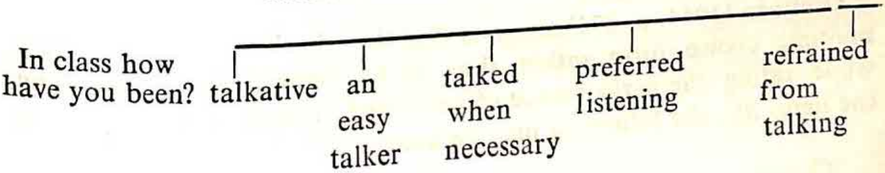
In some numerical scales, the observer or rater is not provided with numbers which he has to use in making judgments. He has to report in terms of descriptive 'cues' and then the researcher assigns numbers to them. For example, while rating performance in a drama, the cues may be the following: very good; good; average; poor; very poor. To these cues the numbers 1 through 5 may be assigned by the researcher.

Numerical rating scales are easiest to construct and to apply. They are simplest in terms of handling the results. However, numerical scales are often rejected in favour of other types of scales because it is believed that they suffer from many biases and errors.

Graphic scales. The graphic scale is the most popular and the most widely used type of rating scale. In this scale a straight line is shown, vertically or horizontally, with various cues to help the rater. The line is either segmented in units or it is continuous. If the line is segmented, the number of parts can be varied. The example of such a scale are illustrated as under:

How effective was the presentation of material aids in the class by the teacher?





There are many advantages of graphic scales. They are simple and easy to administer. Such scales are interesting to the rater and require little added motivation. The raters can fill them out quickly as such scales do not bother raters with numbers. The graphic scale provides opportunity for as fine discrimination as that of which the rater is capable and the fineness of scoring can be as great as desired. As for disadvantages, there is somewhat greater labour of scoring in connection with some formats of graphic scales.

Standard scales. In standard scales a set of standards is presented to the rater. The standards are usually objects of same kind to be rated with pre-established scale values. In its best form, this type is like that of the scales for judging the quality of handwriting. The scales of handwriting provide several standard specimens that have previously been spread over on a common scale by the methods of equal-appearing intervals or pair comparisons. With the help of these standard specimens, a new sample of handwriting can be equated to one of the standards or judged as being between two standards. The man-to-man scale and the portrait-matching scale are other two forms that conform more or less to the principle of standard scales. Man-to-man scale was originally developed for use in connection with military personnel and is of little use elsewhere. The portrait-matching technique was first used in connection with the studies of character by Hartshorne and May (1929).

Rating by cumulated points. The unique and common feature of rating by cumulated points is in the method of scoring. The rating score for an object or individual is the sum or average of the weighted or unweighted points. The 'Check list method' and the 'Guess-who technique' belong to this category of rating.

'Checklist methods' are applicable in the evaluation of the performance of personnel in a job. Hartshorne and May (1929) used this method for evaluating children with respect to character. A list of 80 trait names describing some favourable and unfavourable qualities like cooperative, cruel, thoughtful, humane, greedy, etc., was prepared. Each rater checked every term in the list that he thought applied to a child. The weights of +1 and −1 were assigned to every favourable and unfavourable traits respectively and the child's score was the algebraic sum of the weights.

Guilford (1954, p. 271) suggests that the check list items may be in multiple choice form rather than in true-false form. For example, while rating the performance of personnel in their work assignment, the items like the following may be used:

Cooperate with others
—enthusiastically
—willingly
—indifferently
—grudgingly
—defiantly

His relations with public are:
—outstanding
—creditable
—acceptable
—poor
—detrimental

The 'Guess-who technique' of rating was also developed by Hartshorne and May (1929), for use particularly with child raters. For this purpose some statements in terms of some 'descriptions' like, "Here is one who is always doing little things to make others happy", were constructed and each child was told to list all his classmates who fitted each description, mentioning the same child as many times as necessary. Each child scored a point for each favourable description applied to him, and the total score was the sum total of all such points. Guilford (1954, p. 272) suggests that the scale-values of the 'descriptions' can be found by the method of equal-appearing intervals and the total score for each child is the average of the scale-values of the 'descriptions' that apply to him.

Forced choice ratings. In 'Forced-choice Rating' methods, the rater is asked, not to say whether the ratee has a certain trait or to say how much of a trait the ratee has but to say essentially whether he has more of one trait than another of a pair.

In the construction of a 'Forced-choice rating' instrument descriptions are obtained concerning persons who are recognized as being at the highest and lowest extremes of the performance continuum for the particular group to be rated. Descriptions are analysed into simple behaviour qualities, stated in very short sentences or by trait names, which have been called 'elements' by Sisson (1948) and Guilford (1954, p. 274). These elements are used to construct items and then 'discrimination-value' and 'preference-value' are determined for each 'element'. In forming an item, elements are paired. Two statements or terms with about the same high preference value are paired, one of which is valid and the other not. Two statements or terms with about equally low preference value are also paired, one being valid and the other not.

Two pairs of statements, one pair with high preference value and one with low preference value, are combined in a tetrad to form an item. An example of a tetrad given by Guilford (1954, p. 275) is as under:

——————careless
——————serious-minded
——————energetic
——————snobbish

In the instruction, the rater is asked to react to each tetrad as an item, saying which one of the four best fits the ratee and which one of the four is least appropriate. The tool is tried out in a sample for which there is an outside criterion for the purposes of validating the responses. Then the 'discriminating responses' are determined and 'differential weights' are assigned to each item.

Taylor and Wherry (1951) found that 'leniency error' gets reduced in forced-choice ratings when compared with graphic ratings. However, the research work involved in the construction of forced-choice rating methods is comprehensive and the proponents of the method do not claim complete control of personal biases.

Limitations in Constructing and Using Rating Sales

Rating scales have several limitations some of them are discussed as under:

Constant errors. Ratings based on human judgments are subject to many sources of personal biases or errors. The well-known constant errors in the rating have been classified into six categories by Guilford (1954, pp. 278-280).

1. *The error of leniency.* There is a constant tendency among the raters to rate those whom they know well, or in whom they are ego-involved, higher than they should. Such raters are called 'easy raters'. Some raters become aware of the failing of easy rating and consequently rate individuals lower than they should. Such raters are called 'hard raters'. The leniency error refers to a general and constant tendency for a rater to rate too high or too low for whatever reasons. When rating is too low, the constant error is one of negative leniency. On the other hand, the constant error is one of positive leniency when rating is too high. The positive leniency error is most common and the following arrangement of 'cues' as given by Guilford, (1954, p. 278) on a scale, may prove helpful to counteract it:

Physical Health

↓	↓	↓	↓	↓
poor	fair	good	very good	excellent

In this example only one unfavourable 'cue' is given and most of the range is given to degrees of favourable report. The researcher evidently anticipates a mean rating somewhere near the cue 'good' and a distribution symmetrical about that point.

2. *The error of central tendency.* In the error of central tendency, most of the raters hesitate to rate the individuals on the extremes of the scale and thus tend to rate the individuals on the middle of the scale. It is more common among the individuals who are unknown to the raters. To counteract the error, greater differences in meaning may be introduced between 'cues' near the ends of the scale than between 'cues' near the centre.

3. *The halo-effect.* The error 'Halo-effect' was first mentioned by Wells in 1907 and the error was given its name by Thorndike in 1920. Every rater falls victim to this type of constant error. In the words of Rugg (1921):

> We judge our fellows in terms of a general mental attitude towards the personality as a whole, a like mental attitude towards particular qualities.

Halo-effect is an error which obscures the cluster of traits within an individual. The rater forms a general opinion about the person's merit and his ratings on specific traits are greatly influenced by this general impression. It results in spurious amount of positive correlation between the traits that are rated.

Halo-effect error can never be completely avoided but the rater may take precautions to avoid it to some extent in the rating of a trait where the error is more prevalent. According to Symonds (1925) the error is more prevalent : (*i*) in a trait that is not easily observable; (*ii*) in a trait that is not frequently singled out or discussed; (*iii*) in a trait not clearly defined; (*iv*) in a trait involving reactions with other people; (*v*) in a trait of high moral importance, this involves the so-called traits of character.

Guilford (1954, p. 279) suggests that the practice of rating one trait at a time on all ratees, facilitated by having one trait per page rather than one ratee per page, and the practice of the forced-choice technique may be used to counteract the halo-effect.

4. *The logical error*. The logical error was first pointed out by Newcomb in 1931. It is due to the fact that judges are likely to give similar ratings for traits which they feel logically related to each other. This apparent logical coherence of the traits increases inter-correlation among them. Guilford (1954, p. 279) suggests that this error can be avoided by calling for ratings based on judgments of objectively observable actions rather than abstract and semantically overlapping traits.

5. *The contrast error*. The contrast error was given its name by Murray (1938). The error is due to a tendency for a rater to rate others in the opposite direction from himself in a trait. For example, in a study the raters were asked to rate individuals in the trait of "need for orderliness". It was seen that the raters who themselves were high in orderliness tended to see others as being less orderly than they were, and raters low in orderliness tended to see others as being more orderly than they were. The phenomena of reaction formation and of projection may be responsible for such type of attitudes of raters towards some traits. This error can be avoided to some extent by making the raters feel about the existence of such a phenomena.

6. *The proximity error*. The proximity error was first discovered by Stockford and Bissell in 1949. This error also gives rise to undue co-variances among some traits like the logical error and the contrast error. It has been seen that adjacent traits on a rating form tend to intercorrelate higher than remote ones, their degree of actual similarity being approximately equal. This error can be counteracted to some extent by placing similar traits farther apart and the different ones closer together.

Construction of a Rating Scale

The following points may be kept in view while constructing a rating scale:

1. A trait to be rated should be given a trait name and a definition. It will give the rater a clear conception of the scale or the continuum along which he is to evaluate objects. On the basis of the results of some researches, Guilford (1954, p. 296) has suggested some rules for defining and describing the traits:

(*i*) Traits should be described univocally, objectively, and specifically.

(*ii*) A trait that is to be rated should not be a composite of a number of traits that vary independently.

(*iii*) Each trait should refer to a single type of activity or to the results of a single type of activity.

(*iv*) Traits should be grouped according to the accuracy with which they can be rated.

(*v*) In describing traits, avoid the use of general terms such as 'very', 'extreme', 'average' or 'excellent'.

(*vi*) Traits should be judged on the basis of past or present accomplishments rather than upon what the raters regard as future promise.

(*vii*) In self-ratings, there is no trait in which all individuals overestimate or all underestimate themselves.

(*viii*) Finally, do not use scales for traits on which reliable or more objective data can be obtained. It would be unwise to depend upon ratings of health when medical records are obtainable, or to use ratings on intelligence when mental tests are available.

2. A rating scale should make use of good 'cues'. According to Guilford, (1954, p. 293).

> Rating scale cues have the double purposes of supplementing and reinforcing the definition of the continuum and of providing anchors or mile posts to guide the rater in making quantitative judgments. In order to serve these functions, they must be very carefully written, carefully selected, and carefully placed.

There are several requirements for good cues. Guilford (1954, p. 293) on the basis of a study of cues by Champney (1941), has listed six requirements:

(*i*) *Clarity*. Use short statements, in simple and unambiguous terminology.

(*ii*) *Relevance*. The cue should be consistent with the trait name and its definition as well as with other cues. Avoid bringing into a cue any implications of other traits.

(*iii*) *Precision*. A good cue applies to a point or a very short range on the continuum. There should be no doubt about its rank position among other cues and if possible it should not overlap them in quantitative meaning.

(*iv*) *Variety*. The use of the same terms in all or many of the cues may fail to differentiate them sufficiently. Vary the language used at different scale levels.

(v) *Objectivity.* Cues with implications of good or bad, worthy or unworthy, desirable or undesirable should generally be avoided.

(vi) *Uniqueness.* The cues for each trait should be unique to that trait. Avoid using cues of a very general character, such as 'excellent,' 'superior', 'average', 'poor' and the like.

3. There are no hard and fast rules concerning the number of steps or scale divisions to be used in a rating scale. If the number of steps is too small, the raters are not capable of making much discrimination. On the other hand, too many steps in the scale are beyond the raters' limited power of discrimination. In general, 5 to 7 point scales seem to serve adequately. With motivated, willing, serious and co-operative raters, much finer divisions of the scale prove profitable.

General Advantages of Rating Methods

There are some advantages of rating methods when compared with the method of pair comparisons and method of rank order.

1. Rating methods consume much less time than methods of pair comparisons and rank order.
2. They are far more interesting to the raters, especially if graphic methods are used.
3. Rating methods can be used with raters who have minimum of training.
4. They can be used with large numbers of stimuli.
5. They have much wider range of application and can be used for teacher-ratings, personality ratings, school appraisal, sociological surveys, etc.
6. Best ratings can be obtained by presenting one stimulus to a rater at a time.

Score-card

The score-card is an elaborate form of rating scale. It usually provides for the appraisal of a large number of aspects or characteristics. The presence of each characteristic or aspect is evaluated in terms of a 'numerical point value' and the evaluation of the object observed is done with the help of a 'total weighted score'.

Score cards are commonly used in evaluating schools, textbooks, building sites, libraries, laboratories, worth of any literary or educational institution, etc.

The limitations and problems in constructing the score-card are similar to those of rating scale.

The Opinionnaire or Attitude Scale

The inquiry form that attempts to assess the attitude or belief of an individual is known as an opinionnaire or attitude scale.

Thurstone (1946) defines an attitude as the degree of positive or negative effect associated with some psychological objects. By a psychological object, he means any institution, ideal, symbol, phrase, slogan, job or idea towards which people can differ with respect to positive or negative effect. The concept 'attitude' according to him denotes the "sum total of a man's inclinations and feelings, prejudice or bias, pre-conceived notions, ideas, tears, threats and convictions about any specific topic". The attitude is a personal disposition which impels an individual to react to an object, situation or proposition in favourable or unfavourable ways. For any attitude there is an attitude continuum extending from favourableness through neutral to unfavourableness.

Various 'scaling techniques' have led to the development of different types of attitude scales which provide quick and convenient measure of attitudes. According to Freeman (1965, p. 596) there are three assumptions upon which attitude scales are based:

1. The scale should deal with a controversial question.
2. An individual's feelings and insights in regard to the question will determine his responses to various statements that are made pro and con.
3. The statements can be scaled regarding the degree to which they favour, or are opposed to, the question under consideration.

The 'Method of Equal-Appearing Intervals', and 'Method of Summated Ratings' have been extensively used in attitude or opinion research. The attitude scales which are developed using these scaling techniques consist of a number of carefully edited and selected items called statements. The set of all possible statements that could be made about a given psychological object is called a 'universe of content'. In making the initial list of statements following criteria as listed by Edwards (1957) on the bases of suggestions given by Thurstone and chave (1929), Wang (1932) Bird (1940), Edwards and Kilpatrick (1948) should be kept in view:

1. Avoid statements that refer to the past rather than to the present.
2. Avoid statements that are factual or capable of being interpreted as factual.
3. Avoid statements that may be interpreted in more than one way.
4. Avoid statements that are irrelevant to be psychological object under consideration.
5. Avoid statements that are likely to be endorsed by almost every one or by almost no one.

6. Select statements that are believed to cover the entire range of the effective scale of interest.
7. Keep the language of the statements simple, clear, and direct.
8. Statements should be short, rarely exceeding 20 words.
9. Each statement should contain only one complete thought.
10. Statements containing universals such as *all*, *always*, *none*, and *never* often introduce ambiguity and should be avoided.
11. Words such as *only*, *just*, *merely*, and others of a similar nature should be used with care and moderation in writing statements.
12. Whenever possible, statements should be in the form of simple sentences rather than in the form of compound or complex sentences.
13. Avoid the use of words that may not be understood by those who are to be given the completed scale.
14. Avoid the use of double negatives.

The Method of Equal Appearing Intervals

The method of equal-appearing intervals was originally devised by Thurstone and Chave (1929) and has been modified subsequently in the light of research findings. This technique involves the following steps:

1. *Collecting and editing of statements.* A large number of statements which express varying degrees of intensity of feeling or opinion towards a psychological object are collected. They are collected by making an extensive study of the literature, by consulting experts and research workers. Sometimes the statements are also collected from the essays which are got written from the individuals for whom the attitude scale is to be developed. The statements are edited keeping in view the criteria listed earlier.

2. *The sorting procedure.* The statements are then given to the experts or judges for classification. Thurstone and Chave (1929) used 300 judges for classifying 130 statements they used in constructing a scale for measuring attitude towards church. Subsequent researches by Nystrom (1933), Ferguson (1939), Rosander (1936). Uhrbrock (1934), and Edwards and Kenney (1946) as listed by Edwards (1957, pp. 94-95) indicate that a much smaller number of judges can be used and the number may be as small as 15. Each judge is informed of the purpose of the scale and is asked to classify each statement on an eleven-point scale. This is done by printing each statement on a separate card and judges are then asked to sort each card in one of eleven piles, presumably forming a continuum, according to degree of favourableness or unfavourableness of each statement with respect to the psychological object under study. Instead of sorting the statements into various piles or categories, some researchers have used different methods like rating the statements, etc., to obtain judgements and the

results indicate no significant variations in the ordering of statements from that obtained by following the procedure of sorting as used by Thurstone and Chave. Commenting on the variations in the methods of obtaining judgements, Edwards (1957, p. 98) remarks that the relating ordering of the statements on the psychological continuum is much the same, regardless of which of the methods reported upon is used in obtaining the judgements.

3. *Selection of statements for the final scale.* As a result of the sorting or rating procedure the number of judges for each statement under different categories is found out. Then Q $(=Q_3-Q_1)$ and the median or scale value for each statement are computed. The items for the final scale are selected on the bases of Q and median or scale values, i.e. the items should be fairly and evenly spread on the scale continuum, and Q-values of the selected statements should be as low as possible. Vernon (1962, p. 154) recommends that:

> On a 9-point scale, Q-value should seldom exceed 2.00 and an average of less than 1.50 should be aimed at.

The median or scale values indicate the spread of statements on the scale continuum, and Q or inter-quartile range is a measure of the spread of the middle 50 per cent of the judgements. When there is good agreement among the judges in judging the degree of favourableness or unfavourableness of a statement, Q will be small compared with the value obtained when there is relatively less agreement among the judges. Thurstone and Chave (1929) regarded large Q-value as an indication that a statement is ambiguous.

The final form of the scale is then constructed by selecting 30 to 35 statements which are most relevant, least ambiguous and which cover or represent the different intensities of the attitude. They are then arranged in a random order.

4. *Reliability and validity.* The reliability of the attitude scale developed by following the method of equal appearing intervals is obtained by split-half method. The usual procedure of assigning alternate items to the two forms A and B is slightly modified because that procedure would give one of the two parallel forms a slightly higher mean scale value than the other. This modified procedure was also adopted by Thurstone and Chave, while finding the reliability of attitude scale. In order to make the two forms truly parallel, all the items of the scale are arranged in rank order according to the scale values. Successive pairs are then marked off. The first item in each pair had a slightly lower scale value than the second. In the odd

numbered pairs of the ranked list the first item with the lower scale value is assigned to form A of the scale, and in the even numbered pairs the second item with higher scale value is also assigned to form A of scale. The others are assigned to form B. In this manner two forms are partially identical. Then the attitude scale is administered to a group of subjects who are asked to indicate the statements with which they are in agreement and average scale values for both the forms, A and B, are obtained. The average scale value assigned to form A or to form B is the mean of the scale values of the items which a subject has endorsed in either of the forms. The correlation between the two sets of scores gives the reliability of the half-scale and which is interpreted by means of Spearman-Brown formula for estimating the reliability of the whole scale.

The validity of the attitude scale is obtained by correlating the average attitude scores obtained by finding the mean of the scale values of the endorsed statements with some external criterion. The external criterion may be the performance of the subjects in some related field. The scores of the attitude scale can also be compared with actual behaviour.

The Method of Summated Ratings

The 'method of summated ratings' was introduced by Likert (1932). This method appears to yield similar results to that devised by the method of equal-appearing intervals. The coefficient of correlation between the two types of scales was reported as high as +0.92 in a study by Edwards and Kenney (1946). It has been claimed by Likert (1932), Hall (1934), Rundquist and Sletto (1936), and Edwards and Kenney (1946) that the method of summated rating is simpler, easier and less laborious than that developed by Thurstone. Moreover, it has been found that the time required to construct an equal-appearing interval scale is approximately twice that required by the method of summated ratings. In terms of these advantages the method of summated ratings tends to be used more frequently by the researchers in opinion research. The following steps are involved in the construction of Likert-type scale:

1. *Collecting and editing of statements*. A large number of statements which express an opinion or an feeling towards the psychological object are collected. It is important that they express definite favourableness or unfavourableness to the psychological object. The number of favourable and unfavourable statements should be approximately equal. After the statements have been gathered, they are edited keeping in view the criteria mentioned earlier.

2. *The try out.* The preliminary draft of the scale is administered to a sample of 200 subjects who are selected from the population for which the scale has to be constructed. Each item, or statement, in the attitude scale is followed by five responses, one of which is checked by the subject. The responses, indicating degree of strength of attitude are:

Strongly Agree	(SA)
Agree	(A)
Undecided	(U)
Disagree	(D)
or Strongly Disagree	(SD)

In some situations, 'approve'-'disapprove' are used in place of 'agree'-disagree'. Arbitrary scoring weights of 5, 4, 3, 2, and 1 are used for SA, A, U, D, and SD for the statements favouring a point of view. On the other hand, the scoring weights of 1, 2, 3, 4 and 5 are used for the respective responses for statements opposing this point of view. An individual's score on a particular attitude scale is the sum of his ratings on all items.

3. *Selection of items and preparation of the final draft.* In the method of equal-appearing intervals the items are selected on the basis of Q and scale values. As basis for selecting statements in the method of summated ratings, use is made of some form of item analysis. On the basis of scores based upon the responses to all statements, 25 per cent of the subjects with the highest total scores and 25 per cent of the subjects with the lowest total scores are taken. According to Edwards (1957, p. 152) these two groups provide criterion groups is terms of which to evaluate the individual statements. In evaluating the responses of high and low groups to the individual statements, a ratio of 't' is found out. The value of 't' is a measure of the extent to which a given statement differentiates between the high and low groups. A t-value equal to or greater than 1.75 indicates that the average response of the high and low groups to a statement differs significantly. Finally, 20 to 25 statements with the largest t-values ($t \geqslant 1.75$) are selected for the final draft of the attitude scale.

Other methods of item analysis, such as correlational methods, are also used in evaluating the individual statements instead of the t-test described above. The score for each item is correlated with the total score. Only those items that correlate significantly with the total score are selected. This testing for internal consistency helps in elimina ting statements that are ambiguous or that are not of the same type as the rest of the scale.

4. *Reliability and validity.* The reliability of Likert-type of attitude scale is computed by split-half method. Attempts have been made to validate these types of scale by comparing them with older scales and also by correlating the results with actual behaviour. The former approach has been widely used and a high degree of correlation has been found between certain scales. The latter approach has also been made use of and it has been shown that the predictive value of attitude scale in relation to behaviour is not too high.

Limitations of Attitude Scales

The process of inferring attitude from expressed opinion has many limitations:

1. An individual may conceal his real attitude, and express socially acceptable opinions only.
2. An individual may not really know how he feels about a social issue and he may never have given the idea serious consideration.
3. An individual may not be able to express his attitude towards a situation in abstract. He is able to predict his reaction or behaviour only when confronted with a real situation.

Analysis and Interpretation of Attitude Scores

In equal-appearing interval scales, the attitude score obtained by a single individual has an absolute interpretation in terms of the psychological continuum of scale values of the statements making up the scale. The attitude score of an individual is taken as the mean or median of the scale values of the statements with which he agrees. If this score falls in the middle range of the psychological continuum, the attitude of the individual is described as 'neutral'. If it falls towards the favourable end of the continuum, it is described as 'favourable', and if it falls towards the unfavourable end, it can be described as 'unfavourable'. In an equal-appearing interval type of scale, the interpretation of an attitude score is made independently of the distribution of scores for a particular group of individuals.

On the other hand, the interpretation of an attitude score on a summated-rating scale cannot be made independently of the distribution of scores of some defined or norm group. This is because the summated-rating score corresponding to the 'zero' or 'neutral' point on a favourable-unfavourable continuum is not known as it is known in the case of equal-appearing interval scores. The interpretation of the summated-rating attitude score of an individual in terms of favourableness or unfavourableness is always done with the help of the mean of the norm group.

OBSERVATION

Observation is the process in which one or more persons observe what is occurring in some real-life situation, and they classify and record pertinent happenings according to some planned scheme. It is used to evaluate the overt behaviour of individuals is controlled and uncontrolled situations. Observational methods have occupied an important place in descriptive educational research.

As a good research tool, observation needs proper planning, expert execution, and adequate recording and interpretation.

Planning for Observation

Observation as a research technique must always be expert and directed by a specific purpose. It is neither haphazard nor unplanned. The planning for observation includes definition of specific activities or units of behaviour to be observed, the nature of the groups of the subjects to be observed, the scope of observation-individual or group, determination of the length of each observation period, deciding about the tools to be used in making the observation and recording, etc.

According to Good (1966, pp. 244-245) planning for observation includes the following factors:

1. An appropriate group of subjects to observe.
2. Selection and arrangement of any special condition for the group.
3. Length of each observation period, interval between periods, and number of periods.
4. Physical position of the observer and possible effect on the subject or subjects.
5. Definition of specific activities or units of behaviour to be observed.
6. Entry of frequencies or tallies in the record, as a total for the entire observation period or by sub-division of time within the observation period.
7. Scope of observation, whether for an individual child or for a group.
8. Form of recording, including consideration of mechanical techniques and such quantitative factors as number, time, distance, and spatial relationships.
9. Training of the observer in terms of expertness.
10. Interpretating of observations.

Execution of Observation

An expert execution of observation includes:

1. Proper arrangement of specific conditions for the subject oo

subjects to be observed.

2. Assuming the proper role or physical positions for observing.

3. Focussing attention on the specific activities, or units of behaviour under observation.

4. Handling well the recording instruments to be used.

5. Utilizing the training and experience fairly well in terms of making the observation and recording the facts.

Observation may be either participant or non-participant. In the participant observation the observer becomes more or less one of the group under observation. In such situations the observer will be in sight of the person being observed and may actually take part in some activity with the observed individual or group. The observer may play any one of several roles in observation, with varying degrees of participation, as a visiting stranger, an attentive listener, an eager learner, or a more complete role as participant observer.

In the non-participant observation, observer takes a position where his presence is not disturbing to the group. He may follow in the detail the behaviour of one individual or may describe one or two behaviour characteristics of a dozen or more individuals. In this type of observation, a one-way vision screen that permits the observer to see the subject from seeing the observer, is useful.

Observation may also be classified as unstructured and structured. Unstructured observation is mainly associated with participant observation and it is often such as an exploratory technique. The structured observation are much too formal and they are designed to provide systematic description to test casual hypotheses. Structured observations are executed in controlled situations like classroom or laboratory settings. Interaction analysis of the classroom verbal behaviour of a teacher is an example of structured observation. Structured observations start with relatively specific formulations. There is much less choice with respect to the content of observation. The observer sets up in advance categories of behaviour in terms of which he wishes to analyse the problem, and keeps in mind the time limit under which he has to make the observation. In the unstructured observations, it may not be possible to categorize behaviour in advance of observation. Instead of using predetermined categories, the observer considers aspects of behaviour in terms of their context or the situations of which they are part.

Recording and Interpreting the Observation

The recording of the observation data may either be simultaneous or soon after the observation. In the former case, the observer goes on recording his observation data simultaneously with the occurrence of the phenomena observed. In the latter case, the observer undertakes to record his observations not simultaneously with his actual observation process, but immediately after he has observed for a unit of time while the details are still fresh in his mind.

In viewing, classifying and recording behaviour, the observer must take utmost care to minimize the influence of his biases, attitudes and values on the observation report. The observer should know what he is looking for in a given situation and should carefully and objectively record relevant data. The subjectivity on the part of an observer is partly due to his emotional involvement, his selective perceptions and his different powers of recall. In order to overcome the biases introduced by the human observer, various mechanical instruments are used to obtain a more accurate record of events. The use of cameras, tape-recorders, stop-watch, binoculars, audiometer, stethoscope, light meter, thermometers, one-way vision screen or mirror, etc. allows behaviour to be measured to a degree of accuracy which could not be achieved by the human observer. It is advisable to develop an observation form while making observations. The specific behaviours to be observed and recorded should be listed on this form. The observation form should be simple and the behaviours listed on it may be clearly specified and examples given where necessary.

Observation is done either directly, as when the observer plays a passive role and observes without intervening in any way; or in an interview, where the observer plays a more active role, by asking a series of questions or administering a test, and where he observes the behaviour of the interviewee as well as recording his responses.

In addition to rating scales and checklists, the direct observation make use of: anecdotes, time sampling method, incident-sampling method, and controlled diary method.

1. *Anecdotes.* The anecdote has been the most widely used method for describing naturalistic behaviour. It is a word description of a behaviour episode. There is no set pattern for anecdotal writing, and various styles have been used. Brandt (1972) has given following suggestions for improving the scientific quality of anecdotes:

1. Write an anecdote soon after viewing the incident. If some time lag is necessitated by the situation, jot down a key word or two to aid your memory during the more complete writing.
2. Include the basic action or statements of the chief person in the episode, that is, what he did or said.
3. Include enough setting details to indicate where and when the behaviour occurred, under what conditions, and who are involved.

4. Responses or actions of others to the chief person's behaviour should be included.
5. Use direct quotations wherever possible in order to preserve the flavour of how things were stated.
6. Anecdotes should preserve the sequence of actions and responses of the original behaviour incident.
7. Anecdotes should be objective, accurate, and complete as far as important details are concerned. Good literary style, correct grammar and spelling, and even complete sentences are inconsequential. Words chosen should be precise and unambiguous, nouns and verbs primarily. Subjective terminology exemplified by most adjectives and adverbs, should be used sparingly.
8. If research resources are sufficient, use of tape recorder and typist to transcribe anecdotes into written form generally increases the amount of detail that can be included over simple stenographic or handwritten recording.

Anecdote descriptions have served as useful tools in behavioural research. Such descriptions have been extensively used in examining classroom practices; in analysing community influences on development; in assessing change in interpersonal behaviour of hyperaggressive children undergoing residential treatment.

2. *The time sampling method.* It is not usually possible to observe behaviour continuously for long periods because of the expenses and dynamic nature of behaviour. By 'time-sampling' method, behaviour can be studied for a number of short periods of time which are systematically spaced over the total period of study. For example, a child's behaviour may be studied for the whole school day by means of a series of observation, each of ten minutes duration. It is worthwhile to study children over long periods of time by this technique to get a fair picture of their behaviour.

3. *The incident sampling method.* The incident sampling method concerns the observation of the behaviour of an individual in selected incidents or situations in which he becomes involved. It involves the observation of certain instances of selected behaviour. After a series of such observations of selected behaviour, an observer may have sufficient data to draw inferences about the typical pattern of behaviour of the individual.

4. *The controlled diary method.* The 'cotrolled diary' method was first used by Himmelweit and his associates in 1958 to study the television-viewing behaviour of children and more recently by Wragg in 1968 to investigate the leisure activities of boys and girls. In this method, the observer keeps a diary in which he records certain aspects of the behaviour of a subject. The information gathered by controlled diary method is supplemented by the observation data gathered by other means.

Reliability and Validity of Observational Measurement

In observational measurement, Brandt (1972) mentions about three kinds of reliability. First is revealed by the amount of interobserver agreement in records of the same behaviour. Lack of agreement may indicate insufficient training of the observers, ambiguous identification of characteristics to be rated or described, indistinguishable or overlapping categories, or observations made at somewhat different moments in time.

A second kind of reliability has to do with the inconsistencies of a single observer from one time to another. Estimates of the degree of this kind of reliability can be made by asking as observer to view and rate or code at different times exactly the same behaviour.

A third kind of reliability is concerned with the estimate of the variability of the trait itself. There are significant variations in human behaviour from one time to the next, one situation to another. The estimates of the degree of such variations can be made by collecting a considerable amount of observational data of the same traits and calculate the degree of similarity-dissimilarity over varying types of settings and time periods. Resulting correlation coefficients are often referred to as 'stability coefficients'.

An observational measurement can be valid only to the extent that recorded differences in score represent actual differences in behaviour rather than differences in the impressions made on different observers. A number of extraneous variables, such as maturation of subjects during the course of observation, reactive effects of subjects, change in observers or scores, etc. may produce variations in the obtained measures of observation. To overcome the variations due to the reactive effects it is necessary not to make the subjects feel that they are being observed. An observation should take place in as natural settings as possible and not influenced by the presence of the observer or by his measuring or recording tools.

Best (1977, p. 178) is of the opinion that both reliability and validity of observational measurement are improved when observations are made at frequent intervals by the same observer, or when several observers record their observations independently.

Advantages and Limitations of Observation

If properly used, observation method has the following advantages:

1. It provides a direct procedure for studying various aspects of human behaviour, which may be the only effective way to gather data in a particular situation.

2. It enables the observer to code and record behaviour at the time

of its occurance.

The following are some limitations in the use of observation method:

1. A subject may intentionally attempt to exhibit an artificial behaviour when he knows that he is being observed.
2. Observation method is time consuming and sometimes too costly.

INTERVIEW

The interview is a process of communication or interaction in which the subject or interviewee gives the needed information verbally in a face-to-face situation. Although the interview is generally associated with counselling or psychotherapy, it can be used effectively to collect useful information about individuals in many research situations.

Types of Interview

Interviews may be classified according to the purpose for which they are used and according to their design or structure.

For purposes of research, interview may be used as a tool for gathering data required by the researcher to test hypothesis or solve his problems of historical, experimental, survey or clinical type research. This type of interview is called 'research interview'.

In clinical work, social workers and psychiatrists use interview to secure information about an individual's problem, his past history, job or family adjustment. In such situations the major purposes of interview are diagnosis and treatment. This type of interview is designated as 'clinical interview'.

Interview vary in design or structure. In some situations, an interviewer may interview one individual at one time. It is called an 'individual interview.' In a 'group interview' a group of individuals are interviewed by an interviewer. The size of the group should neither be too small nor too large. According to Good (1966, p. 232):

> The size of the group should not be so large that it is unwiedly or inhibits participation by most members and should not be so small that it lacks substantially greater coverage than in the individual interview. The optimum size is approximately 10 to 12 persons. Social, intellectual, and educational homogeneity are important for effective participation of all group members. A circular seating arrangement, with the interviewer as one of the group, is conductive to full and spontaneous reporting and participation.

In the group interview, the interviewees may hesitate to reveal certain facts or experiences in the presence of others.

Interviews are also classified as 'structured' and 'unstructured'. A structured interview is one in which the procedure to be followed is standardized and is determined in advance of the interview. The same type of questions are presented in the same order to each interviewee, and wording of the instructions to the interviewees is specified. The interviewer has no freedom to rephrase questions, add extra ones, or change the order in which the questions have to be presented. Structured interviews impose a degree of formality which does not permit the interviewer to establish the kind of relationship between himself and the interviewee which is necessary if the interview is to be conducted at some depth.

Unstructured interviews provide greater flexibility. Although the series of questions to be asked and the procedure to be followed are decided upon before hand, the interviewer is largely free to arrange the form and timing of the questions. He can thus rephrase the questions, modify them, and add some new questions to his list. Unstructured interviews are usually labelled as 'focussed', 'depth' and 'non-directive'. The focussed interview aims at some particular event or experience rather than on general lines of inquiry about the event. The depth interview is searching and gives emphasis on psychological and social factors. The non-directive interview permits much freedom to the interviewees to talk about the problem under investigation.

Techniques of Interviewing

Although the interview as a research tool can be modified according to the needs of the research situation, there are some techniques that need to be considered. These techniques deal with preparation for the interview, conduct of the interview, and recording of the data.

1. *Preparation for the interview.* It is necessary to plan for the interview carefully if it is to be effective in obtaining the required information. The interviewer must decide exactly what kind of data the interview should yield, whether the structured or unstructured procedure will be more useful, and how the results of the interview, should be recorded. It is advisable to try out the interview on another person before using it for actual investigation. This may reveal the deficiencies that should be corrected before the actual execution of the interview. The interviewer must have a clear idea of the short of information he needs, and accordingly he may prepare questions which will extract that information.

2. *Conduct for the interview.* In the execution of an interview, a

harmonious relationship between the interviewer and interviewee is most essential. A good rapport helps the interviewee to feel at ease and express himself willingly. In order to establish a good rapport, the interviewer should greet the interviewee in a friendly manner so as to get settled in the new situation in a relaxed manner.

An interviewer generally should start with a pleasant conversation and ask factual non-threatening questions in the beginning. To elicit adequate responses from the interviewee, Turney and Robb (1971, pp. 134-135) have suggested certain rules that should be followed during the conduct of interview to facilitate data collecting:

1. Ask only one question at a time.
2. Repeat a question if necessary.
3. Try to make sure that the subject understands the question.
4. Listen carefully to the subject's answers.
5. Observe the subject's facial expressions, gestures, and tone of voice.
6. Allow the subject sufficient time to answer the question, but do not let the interview drag.
7. Avoid suggesting answers to the questions.
8. Do not show signs of surprise, shock, anger, or other emotions if unexpected answers are given.
9. Maintain a neutral attitude with respect to controversial issues during the interview.
10. Teke note of answers that seem to be vague, ambiguous, or evasive.
11. In the unstructured interview, ask additional questions to follow up clues or to obtain additional information.
12. Use tact and skill in getting the subject back to an area of inquiry when he has stayed too for from the original question.

The interviewer should try to redirect the interview to more fruitful topics when he feels that the required information and data are not emerging. He should wind up the interview before the respondent becomes tired.

3. *Recording of the interview*. Recording the interview is as important as preparation for the interview or conducting of the interview. The interviewer may make use of a schedule, a structured format, rating scale or a tape recorder to record the responses of the interviewee. The use of a tape recorder during the conduct of interview not only eliminates the omissions, distortions, elaborations and other modifications of data usually found in written interviews, but it also provides an objective basis for evaluating the adequacy of the interview data in relation to the performance of the interviewer. The use

of tape recorder also permits the interviewer to devote full attention to the interviewee and save much time of the interviewer which he may have to utilize in recording the responses during or after the interview. The conscious and unconscious selection of the interview data by the interviewer is also avoided in a tape recorded interview. If a tape-recorder is not available, the interviewer may take notes of responses. The notes should include unusual and significant behaviour as well as the responses to questions of the interviewees.

Reliability and Validity of Interview

According to Good (1966, p. 237) the reliability of the data obtained through the interview is affected by the following factors:

1. The desire of many interviewees to make good impressions, particularly in answer to questions relating to generally accepted standards of behaviour.
2. The reluctance of many subjects to reveal highly personal informatian that might appear damaging to the interviewee.
3. An attitude of confidence in and respect for the interviewer, on the part of the interviewee.
4. Content and form of questions, procedures established for the interview, physical setting, mode of recording, accidental distractions, and temporary state of the parties involved in the interview.

For validity of the interview, it is advisable to compare the interviewee's responses with other sources of data.

Advantages and Limitations of the Interview

The interview has certain advantages. It provides an opportunity to the interviewer to question thoroughly certain areas of inquiry. The interview permits greater depth of response which is not possible through any other means. It also enables an interviewer to get information concerning feelings, attitudes or emotions in relation to certain questions.

The interview, however, has certain limitations. It is a time consuming technique. The effectiveness of the interview depends greatly upon the skill of the interviewer not ordinarily possessed by inexperienced researchers. There is a constant danger of subjectivity on the part of the interviewer. The interview is most difficult to employ successfully, for even in the presence of a skilled interviewer some interviewees will not respond freely, frankly and accurately.

SOCIOMETRIC TECHNIQUES

The term Sociometry is defined as the measurement of the social relationships that exists among the members of a group. Sociometric techniques attempt to describe attractions or repulsions between group members by asking them to indicate whom they would select or reject in various situations. Such techniques are used in various educational situations to study social adjustment, group dynamics, learning, motivation, discipline, and other problem areas that involve social relations.

Numerous sociometric techniques are used to measure social interaction within groups, but Sociogram, Sociometric Matrices, Guess-Who Technique and Social Distance Scale are most frequently used in educational research.

Sociogram

The sociogram is often used by the classroom teacher, counsellor, or psychologist to study the interpersonal relationships of groups. In the research situations it is used to study the problems of learning, motivation, discipline, and group dynamics.

To construct a sociogram, the researcher generally begins with the question such as "With whom would you most like to work on a small group project? Give your first, second and third choice." Each member of the group is provided with a form on which he can write the names of the members he has chosen. The group members are assured that the information given by them will be held in strict confidence. If there are boys and girls in the group, then the boys may be represented by triangles and girls by circles. A choice may be represented by a single pointed arrow (→), a mutual choice by an arrow pointing in opposite direction (↔). Rejections may be represented by dotted lines (. . . .). The members who are chosen most often are referred to as 'stars', those not chosen by others as 'isolates'. Small groups made of members who choose one another are 'cliques'.

The next step in the construction of a sociogram is the tabulation of the choices made by the group members. For the sake of illustration, the data giving the name and corresponding number of each member of the group are presented in the following data sheet (Table 5.4). In the first three columns the student's own choices are recorded. In the next three columns the number of times that each student is chosen by others, and in the last column the total choices are recorded. The data are based on the choices, which are taken from the form on which students have listed their preferences.

TABLB 5.4. Sociometric Data Sheet.

Sr. N.	*Student*	*Student choice*			*Times chosen*			*Total choices*
		1st	*2nd*	*3rd*	*1st*	*2nd*	*3rd*	
1	Shyam	2	4	15	0	1	0	1
2	Ali	4	1	3	1	0	1	2
3	Bagat	5	7	2	0	1	3	4
4	Krishana	10	7	3	1	1	1	3
5	Gopal	7	9	10	2	3	1	6
6	Rekha	8	3	5	1	0	1	2
7	Bansi	11	5	3	2	3	0	5
8	Vijay	6	5	9	1	0	0	1
9	Anjani	10	5	6	0	1	2	3
10	Daleep	13	7	12	3	2	1	6
11	Akbar	7	10	4	2	2	0	4
12	Rakesh	5	10	13	0	0	1	1
13	Anand	17	11	19	2	1	3	6
14	Suresh	10	11	13	0	2	1	3
15	Harish	11	16	14	1	0	1	2
16	Abbas	15	14	19	1	1	1	3
17	Satish	19	13	16	1	1	0	2
18	Basant	13	17	9	0	0	0	0
19	Roop	16	14	13	1	0	2	3

The first student on the data sheet is Shyam. In the "students' choices" column this student has selected student number 2 (Ali) as his first choice, student number 4 (Krishana) as his second choice and student number 15 (Harish) as his third choice. The column headed "Times chosen" shows that Shyam is listed as a second choice by one student, but is not listed as a first or third choice by any student. The total number of times he has been chosen by other students, therefore, is one and this is recorded in the "Total choices" column. Similarly the data for other members of group can be interpreted.

From the "Total choices" column it is seen that three students, Gopal, Daleep and Anand, have been chosen six times and one student, Bansi, has been chosen five times. These students are referred to as "Stars". Moreover, one student Basant has not been chosen at all by other group members. Such a student is called an "isolate". Shyam, Vijay and Rakesh seldom have been chosen, and they are called as "neglectees".

The next step is to construct a sociogram. A sociogram provides a diagrammatic picture of the group's relationships. The following procedure is used in constructing the sociogram:

1. Place the names of the stars in the centre of the sociogram, using triangles to denote boys, and circles to denote the girls.

2. Place the names of those receiving the next highest scores (in circles and triangles) near the names in the centre but far enough to permit the drawing of lines connecting triangles and circles.

3. Place the isolates (who receive no scores) and the seldom chosen individuals around the outer area of sociogram.

4. Use lines to indicate choices and arrow heads to show direction of choice. A solid line might be used for a first choice, a broken line for second choice, and a dot-dash line for third choice. Sometimes different colours are also used for first, second and third choices.

Interpretation of the Sociogram

The sociogram is interpreted by noting its general shape and pattern. If the lines indicating choices show that a few individuals have chosen each other but seldom any one else, a clique is indicated. These cliques indicate lack of cooperation. When a segment of the group is completely cut off from the rest of the group, there is evidence of a cleavage or social split in the group. Long lines with much overlapping of choice and much individuals choosing each other indicate a well-integrated group. It means that the group has good leadership, co-operation and group spirit.

The following sociogram, as shown in the Figure 5.3, representing

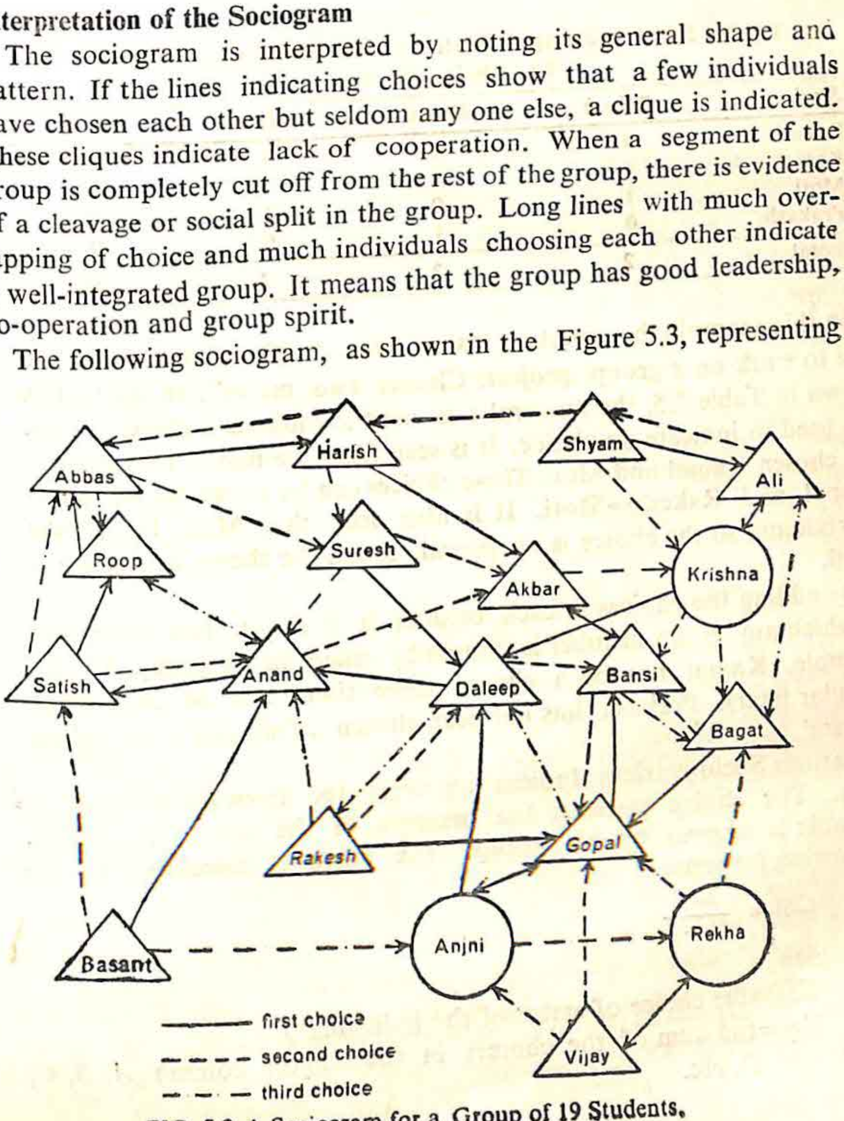


FIG. 5.3 A Sociogram for a Group of 19 Students.

the social relationships of 19 students indicate that there are several stars, a few neglectees and only one isolate. There is no evidence of cliques or cleavages, and the group appears to be well integrated socially.

Sociometric Matrices

A sociometric matrix is a rectangular arrangement of numbers indicating the choices made by the members of a group. An example of a 4×4 matrix (four rows and four columns) is shown in the following Table 5.5.

TABLE 5.5. A Sociometric Matrix of four Member Group for a Two-Choice Question.

Student	*Rakesh*	*Kamal*	*Moti*	*Prakash*
Rakesh	0	1	1	0
Kamal	1	1	0	0
Moti	1	0	1	0
Prakash	0	1	0	0
Total	2	3	2	0

In this example the question asked was, "with whom would you like to work on a group project? Choose two mates". In the matrix shown in Table 5.5, the numeral 1 is used to indicate a choice, while 0 is used to indicate no choice. It is seen from the matrix that Rakesh has chosen Kamal and Moti. These choices can be shown by Rakesh→Kamal and Rakesh→Moti. It is also seen that Moti has chosen Rakesh and so the choice is reciprocal. It can be shown as Rakesh↔Moti.

By adding the choices in each column, it is easy to find the extent to which any group member is chosen by others in the group. In this example, Kamal has been chosen three times and so he is most popular (star). Prakash has not been chosen at all and is termed an 'isolate'.

Various Sociometrices Indices are computed from the sociometric data. The choice status is the measure of the extent to which a member is chosen by his group. The index is calculated by the following formula:

$$CSi = \frac{\Sigma cj}{N-1} \qquad \ldots(5.1)$$

In which

CSj=the choice of status of the individual j

Σcj=the sum of the choices in the subject column A, B, C, D, etc.

N=the number of individuals in the group.

In the above example, the choice status of Kamal is $\frac{3}{3}=1$, of Moti is $\frac{2}{3}=0.66$, and of Prakash is $\frac{0}{3}=0$.

'Social Expansiveness' is another measure. It indicates the extent to which the group members tend to choose others. The formula for computing this index is:

$$E=\frac{\Sigma cij}{N} \qquad \text{...(5.2)}$$

In which

E=index of expansiveness
Σcij=the sum of all choices made by all group members
N=the number of individuals in the group.

Guess-Who Techniques

The Guess-Who technique was developed by Hartshorne and May (1929), for use particularly with child raters. In this technique a student is asked to read each descriptive statement presented to him and then to write down the name of the student who best fits that description. The student may use more than one name for each statement, and he is also allowed to use his own name. Examples of statements that might be used in this kind of technique are:

1. There is a person who is always doing little things to make others happy.
2. There is a person who is disliked by others and has lots of enemies.

The scores for each student is the sum of the number of times he is chosen for each descriptive statement. If the positive statements indicate socially desirable qualities and the negative statements indicate undesirable ones, the total score will be the algebric sum. For example, if a student is mentioned positively 9 times and negatively 3 times, his total score will be $+9+(-3)=+6$. If there are several positive and negative statements for a number of behavioural attributes, it is possible to get a score for each attribute for each student of the group. These scores are useful in the study of individual roles and serve as measure of his reputation.

Social-distance Scale

Social-distance scale is another technique used in the description and measurement of social relationships. This technique was developed by

E.S. Bogardus in 1933. It attempts to assess to what degree an individual, or a group of individuals, is accepted or rejected by another individual or group. Various scaled situations, with score values ranging from acceptance to rejection, are established, and the individual checks his position by choosing one of the points on the scale.

Reliability and Validity of Sociometric Techniques

The concept of reliability and validity may not apply directly to sociometric techniques, but most of the reported studies dealing with the relation of the results obtained by the use of sociometry to actual behaviour show moderately high correlations.

Advantages and Limitations of Sociometric Techniques

Sociometric Techniques are easy to administer and interpret. They are less expensive and less time consuming.

These techniques have certain limitations also. First, information provided is limited by the nature of the sociometric questions and the number of choices used. Secondly, the individuals do not reveal the reasons why they chose or reject some members of the groups. Thirdly, the sociometric data and results do not suggest how the sociometric status of individuals or the social structure of a group can be changed. Finally, the techniques are specific rather than general in their application.

POSTTEST

1. Define the terms population and sample.
2. List and describe the steps involved in the process of sampling.
3. List the various methods of sampling.
4 Define a non-probability sample.
5. List and describe the various types of non-probability sample.
6. Define a probability sample.
7. List and describe the various types of probability sample.
8. Give some examples of various types of non-probability and probability samples.
9. Define the terms statistics and parameter.
10. List and describe the various sampling errors.
11 List the various data gathering tools.
12. Describe the nature and types of psychological tests.
13. What criteria will you use for assessing the worth of a research tool?
14. Define the terms velidity, reliability and usability in the context of psychological tests.
15. List and describe the various types of validity and reliability.
16. Describe the general procedure used in the development of a psychological test.

17. Name and describe in brief some important tests of intelligence, aptitude, creativity, and personality developed or adapted in India.
18. Describe the nature and various types of inquiry forms.
19. Define a questionnaire. Give its various types.
20. List and describe the steps in the construction of a questionnaire.
21. Describe the uses and limitations of a questionnaire.
22. Define a schedule. List and describe the steps involved in the construction of a schedule.
23. Define a checklist. List and describe the steps involved in the construction of a schedule.
24. Describe the uses of (*i*) a schedule and (*ii*) a checklist.
25. Defiine a rating scale. List and describe the various types of rating scale.
26. List and describe the advantages and limitations of rating scales.
27. Define a score card. Describe the advantages and limitations of a score card.
28. Define the term attitude. List and describe some important scaling techniques.
29. Describe the procedure involved in the construction of an attitude scale following the methods of equal appearing intervals and summated ratings.
30. Describe the uses and limitations of attitude scales.
31. Define observation. List and describe the steps involved in the process of observation.
32. Describe the advantages and limitations of observation.
33. Define interview. List and describe various types of interview.
34. List and describe the steps involved in the process of interview.
35. Describe the advantages and limitations of an interview.
36. Define sociometry. List and describe various sociometric techniques.
37. Describe the advantages and limitations of sociometric techniques.

6. Analysis and Interpretation of Data

This chapter reflects the following purposes. First, it deals with the organization of data in terms of editing, classifying and tabulating the information gathered through and administration of various tools on the selected sample or samples. Second, it describes certain modes which are helpful in the analysis of the gathered data. Third, it deals with the various modern computational mechanical aids which are used by researchers in performing complex computations with speed and accuracy. Fourth, it describes brieffy the language and various elements of a computer system. Fifth, it deals with the factors which the researcher should keep in view while making interpretations, and formulating conclusions and generalizations.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Describe the procedures used in the organization of gathered data.
2. List and describe certain modes which are helpful in the analysis of gathered data.
3. List and describe certain modern computational aids which are used by researchers in performing complex computations with speed and accuracy.
4. State what is meant by 'language' of the computer.
5. List and describe briefly the various elements of a computer system.
6. Describe the factors which the researcher should keep in view while making interpretation of the results, and formulating conclusions and generalizations.

The mass of data collected through the administration of various tools on the selected sample or samples in raw. It needs to be analysed and interpreted for drawing sound conclusions and valid generalizations. The first step in the analysis of research data is their organization.

ORGANIZATION OF DATA

Organization of data includes editing, classifying and tabulating the information. Editing implies checking gathered raw data for accuracy, usefulness and completeness. Classification refer to dividing of the data into different categories, classes, groups or heads. For this the researcher may be guided by the nature of the problem, the hypotheses to be verified, or by the responses or characteristics of the sample he has selected. If the problem or hypothesis involved the difference between attitudes of men and women teachers towards co-education at the secondary school stage, the categories male and female serving in government and private aided schools would be clearly indicated. In some situations when the group is sufficiently homogeneous, no breakdown into categories or sub-groups is necessary and it is desirable to describe the group as a whole. However, in the situations where the group is sufficiently heterogeneous, It is desirable to divide the group into homogeneous sub-groups or categories that have in common some distinctive attributes that may be significant for the purpose of the analysis.

Tabulation denotes the recording of the classified data in quantified terms. Quantification not only facilitates their manipulation, but also increases the precision with which they can be analysed. Tabulation is the process of transferring classified data from data-gathering tools to the tabular form in which they may be systematically examined. This process may be performed in a number of ways. In simple and less sophisticated types of research, hand sorting and tabulating procedures are usually employed. More extensive and sophisticated investigations make use of card-tabulating process.

Hand-sorting and Hand-tabulation

Hand-sorting and hand-tabulation require careful planning. It includes method of hand-sorting and recording on tabulation sheets in accurate mathematical terms by marking and counting frequency tallies for different items on which information is sought. The sorting of response sheets in case of psychological tests, questionnaires or opinionnaires in various categories must be done before the tabulation of responses. At times without proper planning a researcher may waste his time and energy by tabulating the responses first and then it might occur to him that it would be interesting to compare the responses of the various sub-groups comprising the sample under investigation. This process would require another handling of the response sheets, questionnaires or opinionnaires and would involve retabulating the responses.

Card Tabulating Process

Several types of cards are used in data tabulation. They include edge marked cards, edge-notched cards, IBM punched cards and mark-sense cards. Edge-marked cards are provided with rows of numbered holes around their four edges. Each numbered space is used to record only one type of information, such as space 3 for sex and space 5 for IQ. These cards can be sorted quickly into groups, but they are suitable only for investigations that involve a small number of variables. On edge-notched cards, data are notched into punched holes spaced along the four edges of the card. For example, hole number 4 is notched if the subject is boy and not notched if the subject is girl. By inserting a long needle through hole 4, the researcher can lift out the cards of the girls and the notched cards of boys will be left. IBM punched cards serve the most efficient and rapid method of processing a large volume of data on cards for electronic computer processing. On mark-sense cards, a subject can mark his answer in the proper space on a card with an electrographic pencil. A machine is used to read these marked data and convert them into punched holes on the same card. The machine further records data on another medium or transmit them directly to a computer.

ANALYSIS OF DATA

Analysis of data means studying the organized material in order to discover inherent facts. The data are studied from as many angles as possible to explore the new facts. Analysis requires an alert, flexible and open-mind. It is worthwhile to prepare a plan of analysis before the actual collection of data. Good, Barr and Scates (1941, pp. 599-601) suggest four helpful modes to get started on analysing the gathered data:

1. To think in terms of significant tables that the data permit.
2. To examine carefully the statement of the problem and earlier analysis and to study the original records of the data.
3. To get away from the data and to think about the problem in layman's terms, or to actually discuss that problem with others.
4. To attack the data by making various statistical calculations.

These exploratory modes may prove very helpful in the analysis of data of any research study and no similarities, differences, trends and significant factors would go unnoticed by the researcher.

Statistical techniques have contributed greatly in gathering, organizing, analysing and interpreting numerical data. The processing of numerical data through statistics calls for competence in the use of statistical methods and for understanding of concepts that underline their development and their application. The researcher must know the strengths and the weaknesses of the statistical methods which he uses so that he may not mislead or be mislead by such methods.

A discussion of two major areas of statistics, descriptive statistics and statistics of inference, is presented in some detail in chapters seven and eight. The main purpose of such discussion is to help the researcher develop an understanding of statistical terminology, and the concepts necessary to study with understanding the literature dealing with educational research. It is also to help the student develop competence and know-how to conduct investigations using simple types of analysis.

MODERN COMPUTATIONAL MECHANICAL AIDS

Modern computational mechanical aids are a boon to the modern researcher. They are used to save time and effort, and to minimize error during organizing and analysis of research data. The increasing and popular use of these computational devices has advanced educational research in terms of both quality and quantity. The computational mechanical aids commonly used are 'desk calculators' and 'digital computers'.

Desk Calculators

The most common type of computational mechanical device available to the researcher is the desk calculator. Its principal advantages are speed and accuracy in performing addition, subtraction, multiplication and division. These operations are performed easily merely by depressing the necessary keys to enter the data and another key to begin the desired operation. The calculations involving combinations of the fundamental operations can also be performed by setting their order as required in computational problem. The desk calculator provides reliable results. At times, improper input of the data or incorrect operation of the machine, or both, furnish erroneous results.

The electro mechanical desk calculators perform the calculations by electrically operated mechanical devices called counters. On the other hand, electronic desk calculators developed recently operate electronically and perform calculations without the use of mechanical counters with greater speed. Some of these electronic calculators are capable of performing operations beyond the four basic operations of addition subtraction, multiplication and division. These additional operations include interpolation, extraction of square roots and reciprocals.

A serious limitation with electro mechanical and electronic desk calculators is that they require a human operator at every stage of operation. However, some of the more modern electronic desk calculators with the help of an attached card reader can accept a series of instructions in the form of "programmes'' and then execute these programmes automatically.

The manufacturers of desk calculators usually provide instruction manuals with them for the use of their operators. These manuals provide directions even for simple operations. If they are studied carefully, the user may not face any difficulty in performing any operation.

Electronic Digital Computers

Electronic digital computers have a long history. The history of their development can be traced back to the first mechanical calculator developed by Blaise Pascal in 1642. But the modern digital computer has grown out of the work of Charles Babbage, a nineteenth century English mathematician. Because of the many errors appearing in scientific tables, Babbage discovered a method of performing complex arithmetic operations quickly, accurately and cheaply. He developed an automatic calculating instrument and later conceived an even more automatic mechine, which used punched cards to control and perform complex operations with a capability for printing the results. Babbage borrowed this idea from the work done in Germany by Muller and from the work done in France by Joseph Jacquard. Late in the nineteenth century Herman Hollerith, a director of the United States Census Bureau, devised a hole punched-card technique for the storage of data, and built a card sorter and card tabulator. The data on the card was transmitted by means of electrical contacts through the punched holes and each hole was reserved for recording a specific item of data. The punched card was a significant development as it is a very important part of modern computer data processing.

During the past thirty years, a stream of ideas for devising faster and accurate methods of processing data has created a computer revolution. In the mid 1940s an electrical impulse computer was developed with circuits using thousands of vacuum tubes. However, these computers were very large and inconvenient to use and required a great deal of space. The heat generated by the vacuum tubes required air conditioning equipment to prevent heat damage. Moreover, the uncertain life of vacuum tubes created mechanical difficulties. During 1950s, the use of electronic switches and replacement of vacuum tubes increased the efficiency of computers. The use of microminiaturized solid-state circuits and a host of other improvements during the 1960s greatly increased the storage capacity, speed, reliability, and range of application of computers. With these developments, modern computers can accept, store and execute more elaborate programmes and can process data four hundred times faster than the earliest one.

LANGUAGE OF THE COMPUTER

A computer is an ingenious and versatile machine. It can manipulate symbols that represent information, has the ability to remember or store these symbols, and can restore them later. However, it cannot determine how to solve a problem. It can only follow direction provided in the form of a "programme". The programme provides instructions about what data to process, how to process the data, and how the results are to be shown.

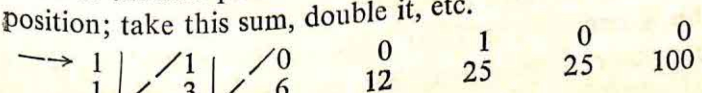
The instructions are written in a language or code that the computer is able to 'read'. A binary code is the most commonly used code within computers. It uses only two digits, 1 and 0, and each of these two digits is called a bit-short for binary digit. The language is important because the electronic parts within a computer can respond only to two states:

(1) current on, or (2) current off;
(1) a switch is opend, or (2) it is closed;
(1) a magnetic material is magnetised in one direction, or (2) in the other.

The binary code-differs from the commonly used decimal system. The decimal system is based on the power of ten; when we move a digit one space to the left and add a zero, its worth is increased ten times. The binary system is based on the power of two; when we move a number to the left, its worth is increased two times.

1,000 100 10 1 ←— Decimal: Successive powers of ten
32 16 8 4 2 1 ←— Binary: Successive powers of two

There are various ways for converting binary number to decimal number. However, there is a short cut method for such conversion which consists of doubling and adding. Starting at the left, copy the value of the first position, double it, and add the value of the next position; take this sum, double it, etc.



The researcher must understand the binary code to know the working of a computer; but he no longer has to use the binary code when writing a programme for the computer. At one time, if he wanted the computer to add something, he might have had to provide the following binary instruction: 101 10101 1101 0001. He may now merely use a "shorthand language" for "add", and the computer will automatically translate the symbol into the correct binary instruction. Today the researcher can learn enough of a shorthand language, such as FORTRAN, to write a programme.

ELEMENTS OF A COMPUTER SYSTEM

The computer system consists of physical equipment, called hardware, and combinations of instruction, called programmes or soft-ware.

Hardware of a Computer System

The essential elements of computer system with regard to hardware are:

1. Input devices,
2. Central processing unit,
3. Main storage, and
4. Output devices.

Figure 6.1 illustrates the functional organization of these elements:

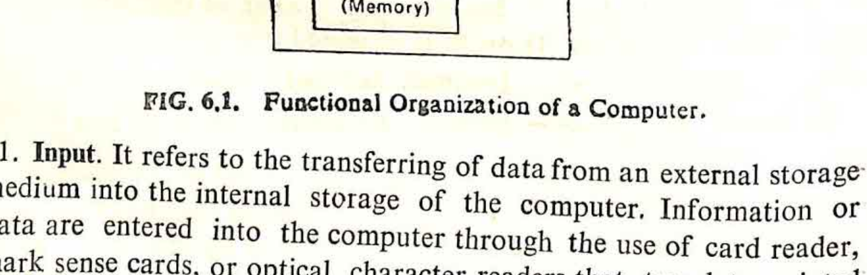


FIG. 6.1. Functional Organization of a Computer.

1. **Input**. It refers to the transferring of data from an external storage medium into the internal storage of the computer. Information or data are entered into the computer through the use of card reader, mark sense cards, or optical character readers that translates printed page information to punched cards or magnetic tape.

Card reader. This is one of the most common input devices. It reads into the computer system data that have been punched into cards. The data are punched into cards by means of a key punch. The punched card is divided into the zone section and digit section. Digits are represented by a single punch in the digit section, while alphabetic and special characters are usually represented by more than one punch in the same column. To facilitate handling and statistical computation, data are usually coded before they are placed on the punch card. The area or space on the punch cards meant for recording a particular data is called 'card field'. The arrangement of data card fields is referred to as the 'card format'. Once the card format has been established for the programme, data can be read quickly and conveniently into the system; and than various manipulations of the data involving summarizing, listing, selecting, etc. with maximum accuracy can be achieved with a minimum of effort.

Mark sense cards. Another device of preparing data for input into a computer system is through the use of mark sense cards. Marks are put in specific locations on a mark sense card. The card is read and the data get automatically punched into the card.

Optical character readers. This is a more quick means of preparing data for computer input. The principal operating action is provided by a rotating drum that transports documents past an optical scanning station. When the pattern of the optically read character agrees with a character pattern in the reader's character-recognition circuits, the character is recorded in input form, that is, punched on cards or recorded on magnetic tape.

Optical character readers can perform an additional function called mark-reading. In this function, marks can be made with an ordinary pencil in specific locations and recorded for input into computer system.

2. **Central processing unit.** This unit controls and directs the entire computer system. It also performs the data treatment functions of addition, subtraction, multiplication, division, shifting, moving, comparing, and storing. The central processing unit consists of the arihmatic or logic unit and the control unit.

The arithmatic or logic unit possesses the capability to make calculations and makes logical decisions about data. At the time of decision making or making calculations, data are sent to the arithmetic or logic unit, processed, and then sent back to be stored for later use or to be used as output.

The control unit supervises and controls all the elements of the system, which must work in perfect coordination. Input data must be entered, stored, and processed as prescribed by the programme. Data must be moved in and out of storage and the input/output devices coordinated by the control unit for maximum efficiency. The control unit performs as it is instructed to perform. It is provided with a series of number codes to initiate and carry out a number of basic operations.

3. **Main storage.** The main storage of the computer system is that storage which is directly accessible to the central processing unit. The data are represented in the main storage by small ferromagnetic rings called 'cores'. The cores are very small in diameter and gets magnetised easily. They retain their magnetic state until it is deliberately changed by reversing the direction of the current. These two opposite magnetic states, as represented by the direction of current, can be used to represent 0 or 1; yes or no; plus or minus; or on or off conditions.

The main storage unit of a computer is like a block af post office boxes. Each box has its own identifying number called an address, and each box contains one fact or one instruction. This address may be assigned by the programmer or automatically by the control unit of the system.

The main storage capacity of the computer system must be considered by the researcher in relation to the size of sample, the number of variables, and the complexity of the calculations involved in the study.

4. **Output devices.** The output unit receives data in machine language from the storage unit and translates in into a language which a human being or a machine can understand. The devices of output are used for intermediate and final output. A card punch, paper tape punch, or magnetic tape unit may be attached to a computer to perform an intermediate output operation. The cards or tapes that are produced for output can in turn be used as input for a later computer operation. The computers make use of printers, graph plotters and visual display equipment for producing output that can be read or interpreted directly by people. The operation of the output device is controlled by the control unit, which operates, according to the instructions entered into the computer system. These instructions specify the medium, nature and format of the output.

Software or Programme of the Computer System

A computer system operates in accordance with specific instructions. Each instruction defines an operation to be performed. It also specifies the data, device, or mechanism needed to carry out the operation. These instructions are referred to as a programme. A computer is unless until a programmer writes a detailed set of instructions to be loaded into its internal storage (memory) unit.

The following are the main six steps involved in writing a programme:

(*i*) To break down the problem into a sequence of elementary steps that the computer can perform.

(*ii*) To make sure that all the necessary steps are included and have been placed in the proper sequence by writing a flow chart, which shows graphically how all parts of the programme fit logically together.

(*iii*) To write the programme in the form of detailed list of instructions on a programme or a coding sheet.

(*iv*) To punch the shorthand language programme on a deck of cards.

(*v*) To translate the shorthand language programme into a machine language programme such as binary language.

(*vi*) To check the major parts of the programme for eliminating any errors.

With the development of shorthand languages and translators like the symbol languages (AU TO CODER and BAL) and the procedure-or problem-oriented languages, (FORTRAN and COBOL), the work of programmers has been greatly simplified. Prepared programmes are now available in the libraries at different computer centres in our country for practically all the statistical techniques that researchers are likely to employ in the analysis of data.

INTERPRETATION OF DATA

Once the research data have been collected and the analysis has been made, the researcher can proceed to the stage of interpreting the results. The process of interpretation is essentially one of stating what the results show. What are their meanings and significance? What is the answer to the original problem? Interpretation is not a routine and mechanical process. It calls for a careful, logical and critical examination of the results obtained after analysis, keeping in view the limitations of the sample chosen, the tools selected and used in the study. There is always an element of subjectivity and the researcher generally commits certain errors while interpreting results of his study.

The researcher should keep the following factors in view while making interpretation of the results.

Ignoring Unstudied Factors in Interpretation

In any type of educational research, the researcher is generally guided by the factors or variables which he has studied during the research process. He totally ignores the influence or effect of unstudied factors while interpreting the results of his study. To totally ignore the unstudied factors and ascribe the findings or results of the research to the occurrence of studied factors alone may be misinterpreting the actual truth, for the findings in any research are conditioned not by one or two but innumerable variables. It is more true in case of experimental or casual-comparative type of research in which the researcher studies a very limited number of variables. For example, if a researcher, who finds that a group of eighth class students following programmed instruction material in social studies has performed better in comparison to another group of students of the same class taught through lecture method, ascribes the better achievement to the method alone and ignores the other possible determining factors like high general mental ability, high achievement motivation, better study habits, interest in the subject and better socio-economic conditions found among the high achieving group, will be misinterpreting the truth.

Ignoring Selective Factors in Interpretation

A researcher may grossly misrepresent the truth if he ignores the selective factors. This is more evident in the studies where a selective group is made the subject of investigation or where a particular factor is operating in the situation studied. For example, if a researcher finds that the boys of a particular tribe are mostly low in intelligence and then concludes, therefore, that the boys of all tribes have a low intelligence, is ignoring the fact that there exist outside the particular tribe, many tribal boys with average or high intelligence. Similarly, to find that in a particular secondary school, the number of the tenth class students failing in mathematics is greater than the number of students failing in other subjects and to conclude from this that mathematics is difficult than other subjects of study is ignoring the fact that the students in mathematics did not receive good instruction in mathematics.

Interpreting Expected Results

While interpreting the expected results the researcher has to keep in mind that he does not go beyond his data support and that he does not forget the limitations of the study. The researcher has to be cautions in reporting all such factors which could account for the results.

Interpreting Negative Results

Researchers often on arriving at results contrary to what they had hypothesized jump to develop a sort of defence mechanism by making a cry of all factors that could have possibly vitiated the results. They often list short comings in terms of the use of inadequate tools or sample fluctuations. These things may be true and there is no harm in reporting all such factor which come in the way of making the study precise. Nevertheless, it is not always correct to get results that confirm hypotheses. At the time of selecting the problem the researcher predicts the expected results of his study in the form of hypotheses on the basis of theory, deduction, and the results of previous research and if the conclusions thus arrived at are very valid then further research becomes pointless.

Hypotheses arise out of guess work and cannot be accepted as right in the beginning without testing them for confirmation. After the research is completed, the researcher is in a position to declare his results with certainty. When the results contradict the original hypothesis of the study, the interpretation and discussion of results should include the researcher's reconsideration of the original hypothesis in the light of his findings. At times researchers are reluctant to interpret and discuss results that contradict the existing known facts. This attitude is not fair and is likely to impede the progress of educational research. It must be noted that the hypotheses are tentative and results can differ from them.

Interpreting Results When the Null Hypothesis is Retained

A retained null hypothesis may occur for several reasons:

(1) There may be no relationship between the variables; or the experimental variable may not be more effective than the control variable.

(2) The null-hypothesis is false, but the interval validity problems of the data contaminated the investigation so badly that the actual relationship beween variables could not be established.

(3) The null-hypothesis is false, but the research design lacked the power to reject it.

Since the researcher is not aware of any of these facts, he could not interpret the results of his study by claiming any one of them as the explanation of his results. Hence, it is not always correct to present a retained null hypothesis as evidence of no relationship between the variables under study.

Interpreting Results When Null Hypothesis is Rejected

If a null hypothesis is rejected and the variables, other than the variables under study, that were likely to affect the results were well controlled in the study, the researcher may say that the behaviour of the variables mentioned in the research hypothesis, which was converted into the null hypothesis for being tested, may be taken to hold good, with a risk of 5 per cent or 1 per cent according to the different levels of significance at which the null hypothesis gets turned down.

Interpreting Unhypothesized Relationships

Any unhypothesized relationship that the researcher may observe during the conduct of a study should not be ignored. It should be recognized and analysed thoroughly. Such results should, however, be viewed with more suspicion than the results based on hypothesized relationships. Unhypothesized relationships have greater possibility of being spurious. Such relationships should be reported but the researcher should consider them as incidental to the main thrust of investigation. They may provide basis of further research.

FORMULATION OF CONCLUSIONS AND GENERALIZATIONS

In the light of the interpretation of data, the researcher has to use all care and caution in formulating his conclusions and generalizations. This final step of research process demands critical and logical thinking in summarizing the findings of the study and compares them with the hypotheses formulated in the beginning. The researcher should not draw conclusions which are inconsistent among themselves or with external realities. The generalizations drawn on the basis of research findings should be in agreement with-facts and should not conflict with the known laws of nature. The suggestions for the application of research findings in practical settings and suggestion of problems for further research may also be provided with conclusions and generalizations.

POSTTEST

1. Describe the procedures used in the organization of gathered data.
2. List and describe certain modes which are helpful in the analysis of gathered data.
3. List and describe certain modern computational aids which are used by researchers in performing complex computations with speed and accuracy.
4. State what is meant by 'language' of the computer.
5. List and describe briefly the various elements of a computer system
6. Describe the factors which the researcher should keep in view while making interpretation of the results, and formulating conclusions and generalizations.

7. Statistical Methods

The organization of this chapter is briefly as follows. The first section deals with definition of the term measurement and its various levels. The second section describes various types of data, the procedures of classifying data and their graphical representation. The third major section lists various statistical measures : measures of central tendency, dispersion, relative positions and relationship. It further illustrates in detail the computation of these measures, the nature and characteristics of normal probability curve, the procedure to consult Normal Table, and applications of normal curve in the field of educational research. The fourth section describes the uses of correlation in the evaluation of reliability and validity of tests, in the technique of factor analysis, and in simple and multiple predictions. The fifth and final section lists and describes briefly the limitation of statistical methods.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives :

1. Define the term measurement.
2. List and describe some important properties of numbers.
3. List and describe the four levels of measurement scales.
4. Define various types of data: parametric and non-parametric.
5. Tabulate a given data into frequency distribution.
6. Illustrate the methods of expressing the class intervals with the help of an example.
7. Compute the cumulative frequencies and cumulative percents for a given frequency distribution.
8. List the four important methods of representing a data graphically.
9. Construct the histogram or column diagram, frequency polygon, cumulative frequency curve, cumulative percentage curve or ogive for a given distribution.
10. List the four descriptive statistical measures.
11. List the three measures of central tendency or averages.
12. Define the mean, the median and the mode.
13. Calculate the mean, the median and the mode for the given (*a*) ungrouped and (*b*) grouped data.
14. Describe the uses of the mean, the median and the mode.
15. List three important measures of spread, dispersion or variability.
16. Define the range, the variance and the standard deviation.
17. Calculate the range for the given ungrouped data.

18. Calculate the variance and the standard deviation for the given (*a*) ungrouped and (*b*) grouped data.
19. Describe the uses of the range and the standard deviation.
20. Describe the nature and characteristics of the normal probability curve.
21. Use the Normal Table.
22. Describe the applications of the normal curve in the field of educational research.
23. List the various measures of comparing individuals on the basis of different types of scores.
24. Define the sigma scores (α scores), the standard scores (Z or T scores), the percentiles and the percentile ranks.
25. Convert a given raw score into a sigma score corresponding to the mean and standard deviation of a distribution.
26. Convert a given raw score into Z score corresponding to the mean and standard deviation of a distribution.
27. Convert a given raw score into T score corresponding to the mean and standard deviation of a distribution.
28. Define a percentile.
29. Calculate certain percentiles for a given distribution of test scores.
30. Define the percentile rank.
31. Calculate the percentile rank of an individual corresponding to his rank in the group to which he belongs.
32. Construct a scattergram for a given bivariate distribution.
33. Plot the regression lines with the help of a given scattergram.
34. List the various measures of relationship.
35. Calculate the product moment correlation for a given (*a*) ungrouped and (*b*) grouped data.
36. Calculate rank order correlation for a given ungrouped data.
37. Calculate biserial correlation and point biserial correlation for a given grouped data.
38. Calculate tetrachoric correlation and phi-coefficient correlation for a given data arranged in a four-fold table.
39. Calculate the coefficient of contingency for a given data.
40. Calculate first order partial correlation for a given data.
41. Calculate multiple correlation for a given data involving three variables.
42. Define correlation ratio.
43. Describe the uses of various methods of correlation.
44. Describe the various uses of correlation.
45. Describe the nature and uses of factor analysis.
46. List certain important procedures of factor analysis.
47. Describe the various limitations of factor analysis.
48. Describe the nature of simple and multiple prediction.
49. Write down the regression equations of (*a*) X on Y and (*b*) Y on X.
50. Define the standard error of estimate.
51. Write down the formula for the standard error of estimate when (*a*) Y measures are predicted from X measures and (*b*) X measures are predicted from Y measures.

52. Predict the X score from the given Y score and calculate the standard error of this prediction.
53. Predict the Y score from the given X score and calculate the standard error of this prediction.
54. Write down the multiple regression equation involving three variable.
55. Write down the formulae for calculating partial regression cyefficients and betas.
56. List and describe briefly the limitations of statistical methods.

Statistical methods are extensively used in educational research. They provide an indispensable tool for collecting, organizing, analysing and interpreting data expressed in numerical terms. By synthesizing the data, these methods can facilitate the derivation of conclusions and formulation of generalizations. Statistical methods use *measurement* as the most precise and universally accepted method for assigning quantitative values to the characteristics or properties of objects or events for the purpose of discovering relationships between variables under study.

MEASUREMENT AND LEVELS

Measurement is defined as the assignment of numbers to objects and events according to logically accepted rules. There are certain properties of numbers that must have parallels in the observed phenomena. These properties are enumerated as under:

1. *The property of identity*. A number has identity. Every number is unique and no other number is exactly the same.

2. *The property of order*. In the number system, numbers have their order or rank i.e. one number being greater than another.

3. *The property of additivity*. In the system of numbers, summing of a certain number with certain other number must yield a unique number. This property is the basis for almost all the more useful operations that can be performed with numbers, because if the numbers can be added, they can also be subtracted multiplied and divided.

It is not necessary that the phenomena to which the numbers are applied must have all the three properties of identity, order, and additivity in order to measure those phenomena. However, the advantage of the numbers used in measurement depends upon how many of the three properties do apply. There are several levels of measurement, the use of which are dependent upon the number of those properties that do apply. A convenient way of classification of the measurement levels is made by Stevens (1951). According to this classification there are four levels of measurement scales.

1. *Nominal scales.* Nominal scales of measurement are used when a set of objects among two or more categories are to be differentiated on the basis of qualitative differences. Usually a number of symbols or numerals are chosen to represent all objects in a given category, thus taking the advantage of the property of identity. We may assign individuals to such categories as sex (males and females), nationality (Indians and Americans), educational level (school students and college students), professional rank (lecturers, readers and professors) etc.

Each individual can be a member of only one category and all the members of the category have the same defined characteristics. Nominal scales are non-orderable and the only arithmatical operation applicable to such scales is counting, the mere enumeration of individuals in each category, class or set. The nominal scale is a primitive form of measurement and the statistical techniques based on counting are permissible in this type of measurement.

2. *Ordinal scales.* The ordinal scales of measurement correspond to quantitative classification of a set of objects. The sets or classes of objects are ordered on some continum in a series ranging from lowest to highest according to the characteristics we wish to measure. The ranking of students in class for height, weight or scholastic achievement are the examples of ordinal scale of measurement. It may be noted that the successive intervals between consecutive points on the ordinal scale may not be equal throughout the entire scale. Suppose we place three students Ram, Sham and Ali, in order of height, Ram tallest, and assign the numbers 3,2, and 1 respectively. All we have is information about serial arrangement. We cannot say that Ram is as much taller than Sham as Sham is taller than Ali, even though the three numbers assigned to them are equally spaced on the scale of measurement. The common arithmetical operations-addition, subtraction, multiplication, and division cannot be legitimately used with ordinal scales, but statistical procedures based on ranks are appropriate.

3. *Interval scales.* If we actually measure the height of the three students Ram, Sham and Ali by using a metre scale and find their heights to be 185 cm, 172 cm and 159 cm., respectively, then we have measurements on a scale of equal units. This scale of measurement is called an interval scale. By this type of measurement we can make some exact and meaningful decisions. We can say that Ram is 13 cm taller than Sham and 26 cm taller than Ali. We can also infer that the

difference in heights between Ram and Sham is twice than that between Sham and Ali.

In the interval scale, the differences between consecutive points on the scale are equal over the entire scale but there is no true zero point on it. The zero point of the scale is chosen conventionally or arbitrarily. Most psychological tests and inventories are based on interval scales. They have no real zero point. It is true that a student may occasionally get a score of zero on a test of arithmetic or general science. However, this does not mean that the student has no knowledge of arithmetic or general science.

We can perform the operations of addition and subtraction on interval scales. The statistical techniques based on these arithmetical operations are permissible. The operations of multiplication and division cannot be legitimately used with interval scales, since these operations presuppose the existence of an exact zero point.

4. *Ratio scales*. The fourth and highest level of measurement is the ratio scale. All the four operations of addition, subtraction, multiplication and division can be used with ratio scales. All statistical techniques are permissible with such scales. Tnese scales have all the characteristics of interval scales, with the additional advantage of a true zero point. It is possible to indicate the complete absence of an attribute. For axample, the zero point on a centimetre scale indicates the complete absence of length or height.

The numerals of the ratio scale can be expressed in ratio relationships. For example, the weight of 4 grams is one-half of 8 grams; the height of 10 cm is one-third of 30 cm, and so on.

Ratio scales are almost non-existent in psychological measurement, except in the area of psychophysical judgement. For example, when we measure reaction time, we use the customary time units, seconds and fractions of a second.

Ratio scales permit us to perform operations of (l) equality of magnitudes, (2) equality of differences, and (3) equality of ratios.

It may be noted that each higher scale can be transformed into a lower category scale but no lower category scale can be transformed into a higher category scale. The statistics permissible under higher category scale would also include the statistics that is permitted under the lower category scale. The higher we go in the hierarchy of the scales, the more relevant information is provided. If the nature of the variables permits, the scale that provides the most precise description should be used.

TYPES OF DATA

Data are figures, ratings observations and other types of information collected in descriptive and experimental studies. In educational research usually two types of data are recognized. They are parametric research usually two types of data are recognized. They are parametric data and non-parametric data.

1. *Parametric data.* Parametric data are got by applying interval or ratio scales of measurement. Data of this type are measured data.

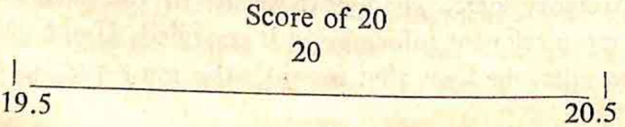
2. *Non-parametric data.* Non-parametric data are got by applying nominal or ordinal scales of measurement. Data of this type are either counted or ranked.

CLASSIFICATION OF DATA

In educational research, while measuring mental and physical traits, we gather data for various attributes or qualities which exhibit differences in magnitude, and which vary along some dimensions. Such attributes or qualities are called variables. Most of the variables in mental and physical measurement fall into continuous series. A continuous series is one which is capable of any degree of sub-division. IQ's, for example, are usually thought of as increasing by increments of one unit along an ability continum which runs from idiot to genius.

Series which exhibit real gaps are called discrete. Gaps in a truly continuous series can be attributed to failure to get enough data, or to defect of the measuring instrument, or to some other factor of the same sort rather than to the lack of measures within gaps.

Test scores or other numbers in continuous series are taken as distances along a continuum rather than as discrete points. A centimetre is the linear magnitude between two divisions on a footrule. Similarly, a score in a mental test is a unit distance between two limits. A score of 20 upon an achievement test, for example, includes all values from 19.5 to 20.5, i.e., any value from a point. 5 unit below 20 to a point .5 above 20. The exact mid-point of this score interval is 20 as shown below.



Other scores are interpreted in the same way. A score of 57 is taken to mean from 56.5 to 57.7, and a score of 160 means from 159.5 to 160.5 Thus, we assume that a score is never a point on the scale but occupies an interval from a half unit below to a half unit above the given number.

Data collected from tests and other types of measuring tools are raw and may have little meaning to the researcher until they have been arranged or classified in some systematic order. The method for classifying scores into what is called a frequency distribution may be illustrated by considering the following scores of eighth grade students on a test of arithmetic.

TABLE 7.1. The Tabulation of Scores on a Test of Arithmetic

2,	5,	8,	6,	10,	9,	3,	2,	1,	6,
7,	5,	4,	5,	0,	2,	5,	6,	7,	4,
3,	4,	6,	8,	5,	7,	4,	3,	5,	3,
4,	9,	5,	7,	6,	3,	1,	4,	5,	6,
0,	4,	5,	9,	10,	7,	2,	3,	3,	1,

It is difficult to see from the above list how the scores are distributed Inspection of scores, however, shows that many scores occur more than once.

We observe that there are two 10s, three 9s, two 8s, five 7s and so on. This suggests that we may arrange the data in columns, as shown in the Table 7.2. In one colume we may arrange the possible marks in descending or ascending order and in the other we record by tallies the number of students who scored these marks. When we have to record five tallies we mark four thus | | and the fifth is made

TABLE 7.2. Frequency Distribution of the Scores on a Test of Arithmetic

Score (*X*)	*Tallies*	*Frequency* (*f*)
10	||	2
9	|||	3
8	||	2
7	𝍸	5
6	𝍸 |	6
5	𝍸 ||||	9
4	𝍸 ||	7
3	|||| ||	7
2	||||	4
1	||.	3
0	||	2
		N=50

across the four thus. | | The tallies are then totalled in the next column. This column thus, lists the number of times each mark occurs. The number of times a particular score value occurs is known as frequency of that score and such an arrangement of the data is known as frequency distribution. The value of the score is generally represented by the symbol *X* and its frequency by the symbol *f*.

In tne Table 7.2 the data have been classified in as many classes as there are score values within the total range of the variables. In the above example, maximum score was 10 and minimum possible score was 0. In this way we had only 11 classes. However, when the range (the interval between the highest and lowest score) of scores is large and we classify in as many classes as there are score vallues, we shall have large number of classes. If the classes are more in number, it becomes difficult to handle them. Therefore, in such cases we reduce the number of classes by arranging the data in arbitrarily defined classes of the variables.

Consider the following scores of 100 college students.

63	78	76	58	95
78	86	80	96	94
46	78	92	86	88
82	101	102	70	50
74	65	73	72	91
103	90	87	74	83
78	75	70	84	98
86	73	85	99	93
103	90	79	81	83
87	86	93	89	76
73	86	82	76	94
95	84	90	73	75
82	86	83	63	56
89	76	81	105	73
73	75	85	74	95
92	83	72	98	110
85	103	81	78	98
80	86	96	78	71
81	84	81	83	92
90	85	85	96	72

Inspection of the above scores, shows that the highest score is 110 and the lowest is 46, so the range is 64 (i.e. 110-46). Therefore, the distribution of scores can be conveniently arranged by dividing the range of 64 into ten or more class intervals if the classes are taken to be of 5 points each and we take starting point 45, then the scores within the range 45 to 49, that is, all scores with the values 44, 46, 47, 48, and 49 will be grouped together to form the lowest class. All scores from 50 to 54, that is, 50, 51, 52, 53, and 54 will form the next class. Similarly, we shall group all scores within the ranges 55 to 59, 60 to 64 and so on. The highest class interval will be 110-114.

TABLE 7.3. Frequency Distribution of the Scores of 100 College Students.

Class intervals	*Tallies*	*Frequency* (*f*)
110-114	|	1
105-109	|	1
100-104	𝍸	5
95-99	𝍸 𝍸	10
90-94	𝍸 𝍸 ||	12
85-89	𝍸 𝍸 𝍸 ||	17
80-84	𝍸 𝍸 𝍸 |||	18
75-79	𝍸 𝍸 ||||	14
70-74	𝍸 𝍸 𝍸	15
65-69	|	1
60-64	||	2
55-59	||	2
50-54	|	1
45-49	|	1
		$N=100$

In Table 7.3, the class intervals have been arranged serially from the smallest at the bottom to the largest at the top. Each class interval covers 5 scores. For each score we have marked tally against the corresponding class interval in which it lies. The first score of 63 is repres entered by a tally placed opposite to class interval 60-64, the second score of 78 by a tally placed opposite to class interval 75-79, and the third score 46 by a tally placed opposite to 45-49. The remaining scores have been tabulated in the same way. When all the 100 scores have been listed, the total number of tallies in each class in the form of frequency s written in the next column *f*. The total of *f* gives the total number of scores (In the present case 100) and is denoted by *N*.

It may be noted that the beginning score of the lowest class interval was taken as 45 and not 46 which was the actual lowest score. Theoretically class intervals of 46-50, 51-55, 56-60, etc. are as good as class intervals of 45-49, 50-54, 55-59, etc, but the second set is easier to handle from the point of view of tabulation as well as computations which will come later.

There are some conventions which may by followed in the selection and formation of class-intervals.

(*i*) The class intervals should be well defined and should not overlap. For example, 0-4, 5-9, etc. should be used in preference to 0-5, 5-10, etc.

(*ii*) The class intervals should be of such a size that between 10 and 20 such class intervals the total range of the observations be covered.

(*iii*) Start the class interval at a value which is a multiple of the size of that interval. For example, with a class interval of 5, the interval should start with the values 5, 10, 15, etc.;with a class interval of 2, the intervals should start with the values 2, 4, 6, etc.

(*iv*) Choose the class intervals of uniform size if possible.

(*v*) A class interval with an odd number of units may be preferable because its mid-point is a whole number rather than a fraction. Since all scores are assumed to fall at the mid-point (for purposes of computing mean), the computation is less complicated.

Methods of Describing the Limits of the Class Intervals in a Frequency Distribution

Table 7.4 illustrates three ways of expressing the limits of the class intervals in the frequency distribution given in the Table 7.3. In (I),

TABLE 7.4. Methods of Describing the Limits of the Class Intervals

Method (I)			*Method (II)*			*Method (III)*		
Class intervals	*Mid point*	*f*	*Class intervals*	*Mid point*	*f*	*Class intervals*	*Mid point*	*f*
110-115	112	1	109.5-114.5	112	1	110-124	112	1
105-110	107	1	104.5-109.5	107	1	105-109	107	1
100-105	102	5	99.5-104.5	102	5	100-104	102	5
95-100	97	10	94.5- 99.5	97	10	95-99	97	10
90-95	92	12	89.5- 94.5	92	12	90-94	92	12
85-90	87	17	84.5- 89.5	87	17	85-89	87	17
80-85	82	18	79.5- 84.5	82	18	80-84	82	18
75-80	77	14	74.5- 79.5	77	14	75-79	77	14
70-75	72	15	69.5- 74.5	72	15	70-74	72	15
65-70	67	1	64.5- 69.5	67	1	65-69	67	1
60-65	62	2	59.5- 64.5	62	2	60-64	62	2
55-60	57	2	54.5- 59.5	57	2	55-59	57	2
50-55	52	1	49.5- 54.5	52	1	50-54	52	1
45-50	47	2	44.5- 49.5	47	1	45 49	47	1
		$N=100$			$N=100$			$N=100$

the interval 45-50 means that all scores from 45 upto but not including 50 fall within this grouping. The intervals in (II) cover the same distances as in I, but the upper and lower limits of each interval are defined more exactly. We have seen that a score of 45 in a continuous series ordinarily means the interval 44.5 upto 45.5; and that a score of 49 means 48.5 upto 49.5. Accordingly, to express precisely the fact that the interval begins with 45 and ends with 49, we may write 45.5 (the begining of score 45) as the lower limit, and 49.5 (end of score49 or beginning of score 50) as the upper limit of this step. The class intervals in (III) express the same facts more clearly than in (I) and less exactly than in (II). Thus 45-49 means that this interval begins with score 45 and ends with score 49; but the precise limits of the interval are not given. The diagram below will illustrate how (I), (II),and (III) are three ways of expressing identically the same facts.

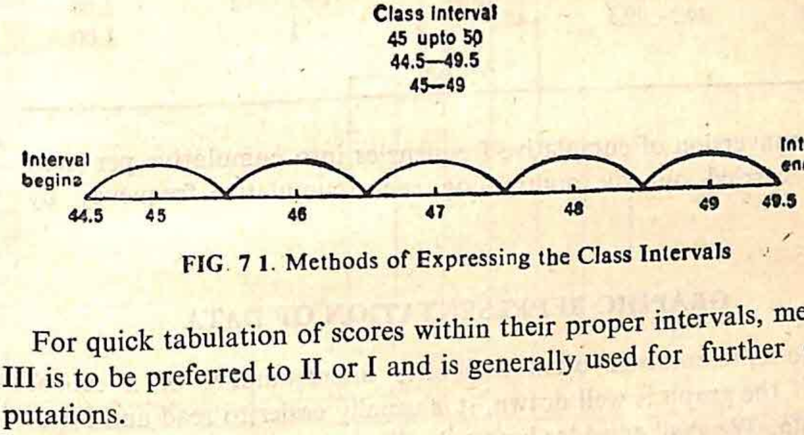


FIG 7 1. Methods of Expressing the Class Intervals

For quick tabulation of scores within their proper intervals, method III is to be preferred to II or I and is generally used for further computations.

Cumulative Frequency Distribution

At times we are not concerned with the frequencies within the class intervals themselves, but rather with number or percentage of values greater than or less than a specified value. This can be obtained by adding successively the individual frequencies. The new frequencies obtained by adding individual frequencies of class intervals are called cumulative frequencies. If we denote individual class frequencies of the successive class intervals by $f_1, f_2, f_3, \ldots f_k$, the cumulative frequencies will be $f_1, f_1+f_2, f_1+f_2+f_3$, and so on.

Table 7.5 shows how from the frequency distribution given in the Table 7.3 we can get cumulative frequencies and cumulative percentages for the distribution.

TABLE 7.5. Cumulative Frequency Distribution of the Test Scores of 100 College Students

Class intervals	*Exact units of class intervals*	*Mid point*	*Frequency* (f)	*Cumulative frequency*	*Comulative percentage frequency*
110-114	109.5-114.5	112	1	99+1=100	100.00
105-109	104.5-109.5	107	1	98+1=99	99.00
100-104	99.5-104.5	102	5	93+5=98	98.00
95-99	94.5- 99.5	97	10	83+10=93	93 00
90-94	89.5- 94.5	92	12	71+12=83	83.00
85-89	84.5- 89.5	87	17	54+17=71	71.00
80-84	79.5- 84.5	82	18	36+18=54	54.00
75-79	74.5- 79.5	77	14	22+14=36	36.00
70-74	69.5- 74.5	72	15	7+15=22	22.00
65-69	64.5- 69 5	67	1	6+1 = 7	7.00
60-64	59.5- 64.5	62	2	4+2 = 6	6.00
55-59	54.5- 59.5	57	2	2+2 = 4	4.00
50-54	49.5- 54.5	52	1	1+1 = 2	2.00
45-49	44.5- 49.5	47	1	1	1.00
			N=100		

The conversion of cumulative frequencies into cumulative per cent can be carried out by multiplying each cumulative frequency by $\frac{100}{N}$.

GRAPHIC REPRESENTATION OF DATA

Graphic representation often facilitates understanding of a set of data. If the graph is well drawn, it is usually easier to read and interpret data. We shall consider here only those graphs which are useful in visualizing the important properties of frequency distribution. These will enable us to comprehend the features of frequency distributions and in comparing one distribution with another. Four methods of graphic representation of data are in general use.

(1) Histogram or column diagram
(2) Frequency polygon
(3) Cumulative frequency curve
(4) Cumulative percentage curve or ogive

The Histogram or Column Diagram

A histogram or column diagram is a graph in which class intervals are represented along horizontal axis called x-axis and their corresponding frequencies are represented by areas in the form of rectangular bars drawn on the intervals.

To prepare a histogram we may proceed as follows:

Step 1. Draw a horizontal line at the bottom of graph paper along which mark off units to represent the class intervals. It is usual to start with the class intervals of lowest value.

Step 2. Draw a vertical line through the extreme end of the horizontal axis along which mark off units to represent the frequencies of the class intervals. Choose a scale which will make the largest frequency (the height) of the polygon approximately 75 per cent of the width of the figure.

Step 3. Draw rectangles with class units as base, such that the areas of rectangles are proportional to the frequencies of the corresponding classes.

The histogram on the data given in Table 7.5 is shown in the Figure 7.2.

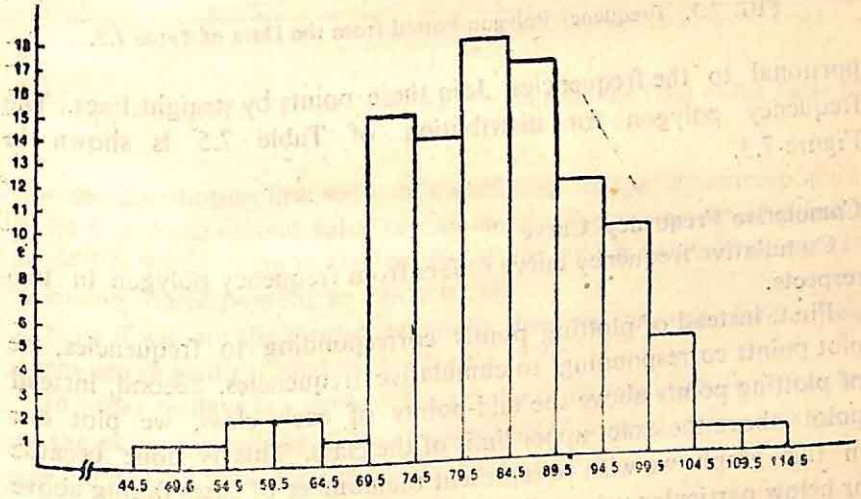


FIG. 7.2. Histogram Plotted from the Data of Table 7.5.

Frequency Polygon

Another method of representing a frequency distribution graphically is, what is known as the frequcncy polygon. The first two steps are identical with those used in the construction of the histogram. The third step to be followed is given as under:

Step 4. Directly above the point on the horizotal axis representing the mid-point of each class interval, plot the points at a height pro-

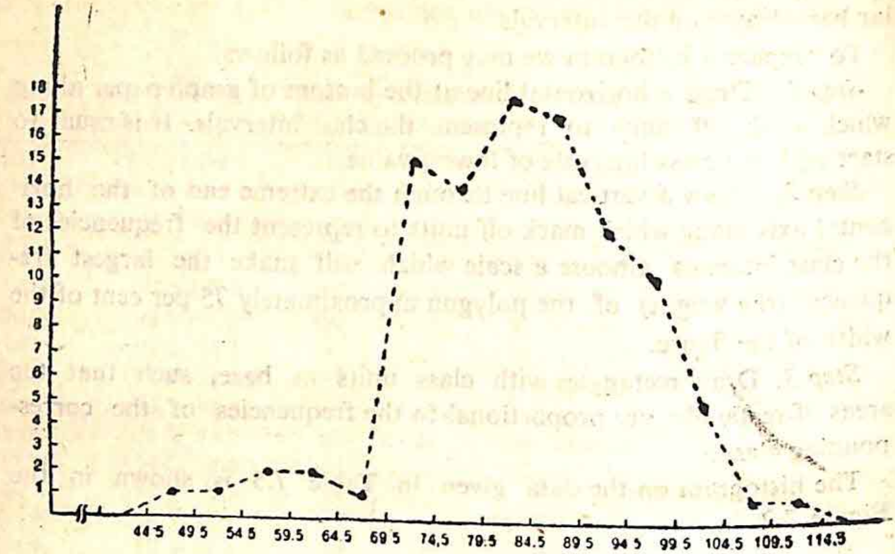


FIG. 7.3. Frequency Polygon Potted from the Data of Table 7.5.

portional to the frequencies. Join these points by straight lines. The frequcncy polygon for distribution of Table 7.5 is shown in Figure 7.3.

Cumulative Frequency Curve

Cumulative frequency curve differs from frequency polygon in two respects.

First, instead of plotting points corresponding to frequencies, we plot points corresponding to cumulative frequeneies. Second, instead of plotting points above the mid-points of each class, we plot our points above the exact upper limit of the class. This is done because in this graph we wish to represent the number of cases falling above or below particular values. For understandig the method of drawing this graph let us consider the frequency distribution given in the Table 7.5.

As illustrated in the column 5 of Table 7.5 the first cumulative frequency from the lower end the distribution (i.e. of of class interval 44.5.49.5) is 1, the next cumulative frequency of class interval 49.5-54.5 is 2, and so on. The last cumulative frequency is equal to 100, or N, the total frequency.

For drawing cumulative frequency curve each cumulative frequency is plotted corresponding to the exact upper limit of the class interval.

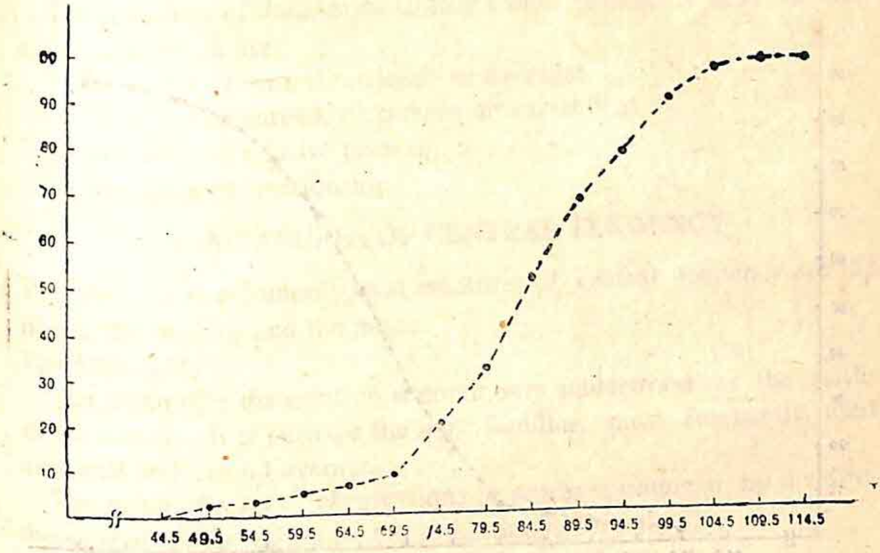


FIG. 7.4. Cumulative Frequency Curve Plotted from the Data of the Table 7.5.

For the distribution first value of cumulative frequency corresponding to 49.5 is 1; the second value of cumulative frequency corresponding to 54.5 is 2 and so on to the last point for which value of cumulative frequency corresponding to 114.5 is 100.

Now if we join the plotted points by lines we get cumulative frequency curve as illustrated in the Figure 7.4.

In order to have the curve being on the horizontal axis it is started at the exact lower limit of the lower class (44.5 in the present case), the cumulative frequency of which is taken to be 0.

The Cumulative Percentage Curve or Ogive

The cumulative percentage curve or ogive differs from the cumulative frequency curve in that frequencies are expressed as cumulative per cents of *N* on the vertical axis instead of as cumulative frequencies. In the column (6) of the Table 7.5, the cumulative frequencies are expressed as percentages of N. After finding cumulative percentage frequencies we may then plot these frequencies corresponding to upper exact limits of class intervals. A curve joining the points thus obtained is called cumulative percentage curve or ogive.

The cumulative percentage curve or ogive of the distribution represented in the Table 7.5 is illustrated in the Figure 7.5.

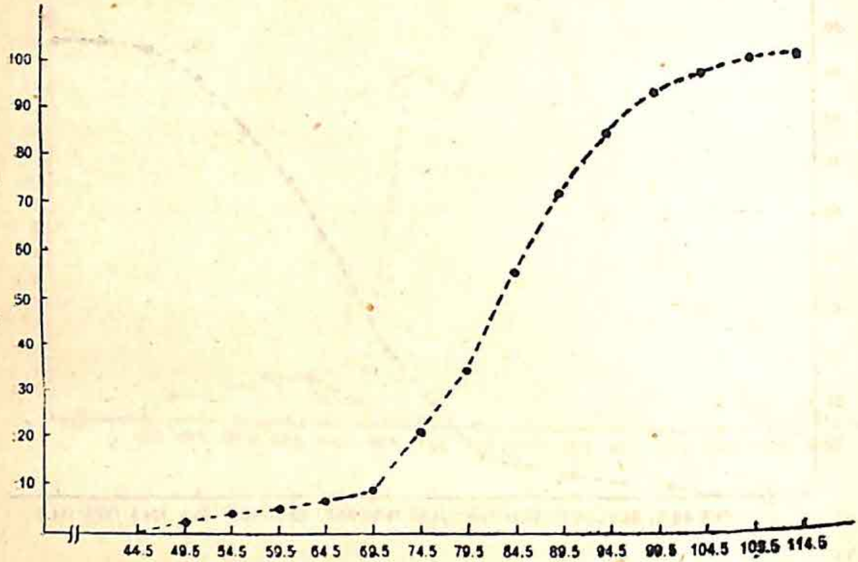


FIG. 7.5. Cumulative Percentage Curve or Ogive Plotted from the Data of the Table 7.5.

Uses of Ogive

1. Percentiles and percentile ranks may be determined quickly and fairly accurately from the ogive when the curve is carefully drawn and the scale divisions are precisely marked.

2. A useful overall comparison of two or more groups is provided when ogives representing their scores are plotted upon the same horizontal and vertical axes.

3. Percentile norms are determined directly from ogives.

DESCRIPTIVE STATISTICAL MEASURES

Descriptive statistical measures are used to describe the characteristics of the sample or population in totality. They limit generalization to the particular group of individuals observed or studied. No conclusions are extended beyond this group. The statistical analysis based on the computation of descriptive statistical measures is mostly applied in action research, and provide valuable information about the nature of a particular group and that group only.

The measures of descriptive statistics most commonly used in educational research are:

1. Measures of central tendency or averages.
2. Measures of spread, dispersion or variability.
3. Measures of relative position.
4. Measures of relationship.

MEASURES OF CENTRAL TENDENCY

The three most commonly used measures of central tendency are the mean, the median and the mode.

The Mean (M)

The mean of a distribution is commonly understood as the arithmetic average. It is perhaps the most familiar, most frequently used and well understood average.

The mean of a set of observations or scores is obtained by dividing the sum of all the values by the total number of values.

Mean for Ungrouped Data

The formula for finding the mean for ungrouped data is

$$M=\frac{\Sigma X}{N} \tag{7.1}$$

in which

$M=$ mean
$\Sigma=$sum of
$X=$scores in a distribution
$N=$total number of scores.

To illustrate the use of this formula, let us consider the data:

42, 46, 30, 37, 29, 41, 50, 31, 39, 43

Using the formula (7.1)

$$M=\frac{\Sigma X}{N}=\frac{42+46+30+37+29+41+50+31+39+43}{10}$$

$$=\frac{388}{10}$$

Mean for Grouped Data

When the number of observations of scores is large, we first group the data in a suitable frequency distribution and then compute the mean using the formula:

$$M=\text{A.M.}+\frac{\Sigma fx'}{N}\times i \tag{7.2}$$

in which

$M=$mean
A.M.=assumed mean

TABLE 7.6 Frequency Distribution of Test Scores of 40 Students

Scores	*f*
70-71	2
68-69	2
66-67	3
64-65	4
62-63	6
60-61	7
58-59	5
56-57	4
54-55	3
52-53	3
50-51	1
	$N=40$

Σ=sum of

f=frequency of the class interval

x'=deviation of the score from the assumed mean divided by length of the class interval

i=width of the class interval

N=total number of scores.

To illustrate the use of the formula (7.2), consider the grouped frequency distribution given in Table 7.6.

Computations

Step 1. Put the class intervals in exacts limits

TABLE 7.7. The Calculation of Mean from Data Grouped into a Frequency Distribution

Scores	*Mid-point (X)*	*f*	*x'*	*fx'*
69.5-71.5	70.5	2	5	10
67.5-69.5	68.5	2	4	8
65.5-67.5	66.5	3	3	9
63.5-65.5	64.5	4	2	8
61.5-63.5	62.5	6	1	6
59.5-61.5	$\boxed{60.5}$ A.M.	7	0	0
57.5-59.5	58.5	5	—1	— 5
55.5-57.5	56.5	4	—2	— 8
53.5-55.5	54.5	3	—3	— 9
51.5-53.5	52.5	3	- 4	—12
49.5-51.5	50.5	1	—5	— 5
	$N=40$			$\Sigma fx'=2$

Step 2. Find the mid-point of each class interval and take the assumed mean at the interval which has the maximum frequency.

Step 3. Find the deviation (x') of each mid-point score from the assumed mean and divide it by the length of the class interval.

Step 4. Compute fx' for each class interval.

Step 5. Find the sum of all fx'

Step 6. Apply the formula (7.2) to compute the mean.

Using the formula (7.2)

$$M = \text{A.M.} + \frac{\Sigma fx'}{\text{N}} \times i$$

$$60.50 + \frac{2}{40} \times 2$$

$$= 60.50 + 0.10$$

$$= 60.60$$

The Median (Mdn)

The Median is a point in an array, above and below which one-half of the scores fall. It is a measure of position rather than of magnitude.

Median for Ungrouped Data

If the scores are ungrouped and their number is small, we arrange the scores in order of magnitude and determine the middle score by counting up half the value of N (the number of scores). When the number of scores (N) is odd, the mid score is the median. For example, 36 is the median of the scores:

31, 33, 36, 37, 40

When the number of scores (N) is even, the median is the mid-point between two middle scores.

For example, $\frac{(25+27)}{2} = 26$ is the median of the scores:

19, 21, 25, 27, 30, 34

Median for Grouped Data

In case of a grouped data, we first prepare the cumulative frequency distribution and median is calculated using the formula:

$$\text{Mdn} = l + \frac{(\text{N}/2 - F) \times i}{f} \tag{7.3}$$

in which

l=exact lower limit of the class interval upon which the median lies

i=width of class interval in which the median falls

f=frequency within the class interval upon which the median lies

F=Sum of all the frequencies below l

$\frac{N}{2}$=one-half the total number of scores.

To illustrate the use of formula 7.3, consider the data of the Table 7.8:

Here $\frac{N}{2}=20$; $l=59.50$; $F=16$; $f=7$; and $i=2$

Using the formula (7.3)

$$\text{Median}=l+\frac{(N/2-F)}{f}\times i$$

$$=59.50+\frac{(20-16)}{7}\times 2$$

$$=59.50+1.14$$

$$=60.64$$

TABLE 7.8. The Calculation of the Median from Data Grouped into a Frequency Distribution

	Score	*f*	*F*
	69.5-71.5	2	40
	67.5-69.5	2	38
	65.5-67.5	3	36
	63.5-65.5	4	33
	61.5-63.5	6	29
Median class	59.5-61.5	7	23
	57.5-59.5	5	16
	55.5-57.5	4	11
	53.5-55.5	3	7
	51.5-53.5	3	4
	49.5-51.5	1	1
		$N=40$	

The Mode

The mode is defined as the most frequently occurring score in a distribution. If there is only one value which occurs a maximum number of times, then the distribution is said to have one mode or to be unimodel.

In some distributions there may be more than one mode. A two-mode distribution is bimodal; more than two, multimodal.

Mode for Ungrouped Data

In a simple ungrouped series of measures or scores the crude or empirical mode is that single measure or score which occurs most frequently. For example, in the series 9, 10, 11, 16, 18, 18, 19 and 21, the most often recurring measure, namely, 18, is the crude or empirical mode.

Mode for Grouped Data

When data are grouped into a frequency distribution, the crude mode is usually taken to be the mid-point of that interval which contains the largest frequency. In the example given in Table 7.6, the interval 60-61 contains the largest frequency and hence 60.5, its mid-point, is the crude mode.

When calculating the mode from a grouped data represented in a frequency distribution, we distinguish between the true mode and crude mode. The true mode is the point of greatest concentration in the distribution; that is, the point at which more measures fall than at any other point. The formula for computing the true mode is

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i \tag{7.4}$$

in which

l=lower limit of modal class *i.e.* the class interval having maximum frequency

f_m=frequency of modal class

f_1=frequency of class interval preceding the modal class

f_2=frequency of class interval following the modal class

i=width of the modal class.

Let us illustrate the use of this formula to the data given in the Table 7.7. Clearly, maximum frequency is 7 which lies in the class interval 59.5-61.5.

Therefore, modal class is 59.5-61.5

Here $f_m=7, f_1=5, f_2=6, i=2$ and $l=59.5$

Using the formula (7.4)

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$=59.5 + \frac{7-5}{14-5-6} \times 2$$

$$=59.5 + \frac{2}{3} \times 2$$

$$=60.83$$

Selection of Measures of Central Tendency

Generally, the mean will be a more useful measure of central tendency than the median and mode.

The mean is an appropriate statistics when

1. the scores are distributed symmetrically about the centre of the distribution;
2. the most stable measure of central tendency is desired;
3. additional statistics are to be computed later;
4. the centre of gravity of a sample is desired.

Compute the median when

1. the exact mid-point of the distribution is wanted;
2. there are extreme scores which would markedly affect the mean. Extreme scores do not disturb the median. For example, in the series of scores 6, 7, 8, 9 and 10, both mean and median are 8. But if 10 is replaced by 31, the other scores remaining the same, the median is still 8 but the mean is 12.2
3. there is not sufficient time to compute a mean;
4. we are interested in whether cases fall within the upper or lower halves of the distribution and not particularly in how far they are from the central point;
5. an incomplete distribution is given.

The Mode is used when

1. the quickest estimate of central tendency is wanted;
2. a very rough estimate of central tendency will suffice;
3. we wish to know the most typical case.

MEAUSRES OF VARIABILITY

Although measures of central tendency are very useful in describing the nature of a distribution of measures, they will not give the researcher a complete picture of the data. These statistics will not tell the researcher how the scores tend to be distributed. For this another kind of statistic, the measure of variability is used. It is also called the measure of spread or dispersion.

There are several measures of variability, but we shall limit our discussion to those that are most useful to a researcher. These include the range, the variance and the standard deviation.

The Range

The range is the most general and simplest measure of variability. It is defined as the difference between the most extreme scores in a distribution. Suppose the scores of a group of 10 students are:

100, 90, 82, 81, 80, 80, 79, 78, 70, 60

The range for this distribution will be (100-60)=40. Although it has the advantage of being easily calculated, the range has some serious limitations as a measure of variability. First, the value of range is based on only two extreme scores in the total distribution, which means that it does not give any idea of the variation of many other scores of the distribution. Second, it is not a stable statistic, because its value can differ from sample to sample drawn from the same population.

The Variance and the Standard Deviation

The average of the squared deviations of the measures or scores from their mean is known as the variance. The standard deviation is the positive square root of variance.

The Variance and Standard Deviation for the Ungrouped Data

The variance for ungrouped data is found by using the formula:

$$\sigma^2 = \frac{\Sigma x^2}{N} \qquad (7.5)$$

in which

σ^2=variance of the sample

x=deviation of the raw score from the mean

N=number of scores or measures.

To illustrate the use of this formula, let us consider the data:

22 20 19 17 17 17 16 16 14 12

Since we will need the deviation of each score from mean, or average score, we will first compute the mean.

Using formula (7.1)

$$\text{Mean}=M=\frac{\Sigma X}{N}=\frac{170}{10}=17.$$

Now we subtract the mean from each raw score in the distribution.

TABLE 7.9. Distribution of Test Scores of Ten Students

Score (X)	*Deviation (X-M)* x	*Deviation Squared* x^2
22	5	25
20	3	9
19	2	4
17	0	0
17	0	0
17	0	0
16	—1	1
16	—1	1
14	—3	9
12	—5	25
		$\Sigma x^2=74$

Using formula (7.5)

$$\sigma^2=\frac{\Sigma x^2}{N}$$
$$=\frac{74}{10}=7.4$$

Now, since we know the variance. σ^2, we need only the positive square root of it to get the standard deviation.

$$\text{Standard deviation}=\sigma=\sqrt{\frac{\Sigma x^2}{N}} \quad (7.5a)$$
$$=\sqrt{7.4}$$
$$=2.72$$

A much less complicated method uses the raw scores instead of the deviation scores. The raw-score formulae for variance and standard deviation are given as

$$\text{Variance}=\sigma^2\frac{N\Sigma X^2-(\Sigma X)^2}{N^2} \quad (7.6)$$

$$\text{Standard deviation}=\sigma=\frac{\sqrt{N\Sigma X^2-(\Sigma X)^2}}{N} \quad (7.7)$$

in which

X=raw score

N=the number of scores in the distribution.

Using the same set of data, we can calculate the variance and standard deviation with the raw-score formulae.

TABLE 7.10. The Calculation of Variance and Standard Deviation from Original Scores when the Assumed Mean is taken at Zero and Data are Ungrouped

Score X	X^2
22	484
20	400
19	361
17	289
17	289
17	289
16	256
16	256
14	196
12	144
$\Sigma x=170$	$\Sigma X^2=2964$

Using formula (7.6)

$$\text{Variance}=\sigma^2=\frac{N\Sigma X^2-(\Sigma X)^2}{N^2}$$

$$=\frac{10\times 2964-(170)^2}{100}$$

$$=\frac{29640-28900}{100}$$

$$=\frac{740}{100}=7.4$$

Using the formula (7.7)

$$\text{Standard Deviation}=\sigma=\frac{\sqrt{N\Sigma X^2-(\Sigma X)^2}}{N}$$

$$=\frac{\sqrt{10\times 2964-(170)^2}}{10}$$

$$=\frac{27.2}{10}$$

$$=2.72$$

Variance and Standard Deviation for Grouped Data

When the number of observations or scores is large, we first group the data in a suitable frequency distribution and then compute the variance and standard deviation using the formulae.

$$\text{Variance}=\sigma^2=\frac{i^2}{N^2}\left[N\Sigma fx'^2-(\Sigma fx')^2\right] \quad (7.8)$$

$$\text{Standard deviation}=\sigma=\frac{i}{N}\sqrt{N\Sigma fx'^2-(\Sigma fx')^2} \quad (7.9)$$

in which

i=width of the class interval
N=total number of scores
f=frequency of the class-interval
x'=deviation of the raw-score from the assumed mean divided by the length of class interval.

To illustrate the use of these formulae let us consider the distribution given in Table 7.11.

Using the formula (7.8)

$$\text{Variance}=\sigma^2=\frac{i^2}{N^2}\left[N\Sigma fx'^2-(\Sigma fx')^2\right]$$

$$=\frac{(5)^2}{(54)^2}\left[(54\times 247)-(-41)^2\right]$$

$$=99.94$$

TABLE 7.11 **The Calculation of Variance and Standard deviation from Data Grouped into a Frequency Disiribution**

Class interval	*f*	*X*	*x'*	*fx'*	*fx'*^{2}
93-97	1	95	4	4	16
88-92	3	90	3	9	27
83-87	2	85	2	4	8
78-82	7	80	1	7	7
73-77	12	75	0	0	0
68-72	10	70	—1	—10	10
63-67	9	65	—2	−18	36
58-62	5	60	−3	—15	45
53-57	3	55	—4	− 12	48
48-52	2	50	—5	−10	50
	N=54			$\Sigma fx'=-41$	$\Sigma fx'^2=247$

Using the formula (7.9)

$$\text{Standard Deviation} = \sigma = \frac{i}{N}\sqrt{N\Sigma fx'^2 - (\Sigma fx')^2}$$

$$= \frac{5}{54}\sqrt{54\times 247 - (-41)^2}$$

$$= 9.99$$

Selection of Measures of Variability

Generally, the standard deviation is a more useful measure of variability.

Range is used when

1. the data are too scant or too scattered to justify the computation of a more precise measure of variability;
2. a knowledge of extreme scores or of total spread is all that is wanted.

Standard deviation is used when

1. the statistic having the greatest stability is sought;
2. extreme deviations exercise a proportionally greater effect upon the variability.
3. other statistics are subsequently to be computed.

The standard deviation is a very useful device for comparing characteristics that may be quite different or that may be expressed in different units of measurement. It serves as a means of describing status or position of an individual in a group. Before this concept is developed further, it is essential to understand the nature of the normal probability curve and its application in the educational research.

NORMAL PROBABILITY CURVE

The normal frequency distribution curve is based upon the law of probability or the probable occurrence of certain events. The 'probability' of a given event is defined as the expected frequency of occurrence of this event among events of a like sort. It may be stated mathematically as a ratio. The probability of an unbiased coin falling heads is $\frac{1}{2}$, and probability of a dice showing a four-spot is $\frac{1}{6}$, These ratios, called 'probability ratios' are defined by that fraction, the numerator of which equals the desired outcome or outcomes, and the denominator of which equals the total possible outcomes. A probability ratio always falls between the limits .00 (impossibility of occurrence) and 1.00 (certainty of occurrence). All possible degress of likelihood may be expressed by appropriate ratios between these limits.

If a fair coin is tossed either a head (H) or a tail (T) will turn up. The probability that a head will appear is one chance in two. Expressed as a ratio, therefore, the probability of H is $\frac{1}{2}$; of T is $\frac{1}{2}$; and we have:

$$(H+T)^1=H+T$$
$$=\tfrac{1}{2}+\tfrac{1}{2}=1.00$$

If we toss two coins, C_1 and C_2, at the same time, there are four possible arrangements which the coins may take:

$$\begin{array}{cccc} (1) & (2) & (3) & (4) \\ C_1\ \ C_1 & C_1\ \ C_2 & C_1\ \ C_2 & C_1\ \ C_2 \\ H\ \ H & H\ \ T & T\ \ H & T\ \ T \end{array}$$

Both coins (C_1) and (C_2) may fall H; (C_1) may fall H and (C_2) T; (C_1) may fall T and (C_2) H; or both coins may fall T. Expressed as ratios, the probability of two heads is $\frac{1}{4}$ and the probability of two tails is $\frac{1}{4}$. Also, the probability of an HT combination is $\frac{1}{4}$, and of a TH combination is $\frac{1}{4}$; and since it ordinarily makes no difference which coin falls H or which falls T, we may add these two ratios to obtain $\frac{1}{2}$ as the probability of an HT combination. The snm of these probability ratios is $\frac{1}{4}+\frac{1}{2}+\frac{1}{4}=\frac{4}{4}=1$

The expected appearance of heads and tails can be expressed as:

$$(H+T)^2=H^2+2HT+T^2$$

in which

1 H^2=both heads: 1 out of 4; probability ratio $=\frac{1}{4}$
2 HT=1 head 1 tail : 2 out of 4; probability ratio$=\frac{1}{2}$
1 T^2=both tails : 1 out of 4; probability ratio $=\frac{1}{4}$
Total=4 Total=1

If we toss three coins (C_1), (C_2) and (C_3) simultaneously, there are eight possible outcomes:

(1) $C_1\ C_2\ C_3$ $H\ H\ H$	(2) $C_1\ C_2\ C_3$ $H\ H\ T$	(3) $C_1\ C_2\ C_3$ $H\ T\ H$	(4) $C_1\ C_2\ C_3$ $T\ H\ H$
(5) $C_1\ C_2\ C_3$ $T\ H\ T$	(6) $C_1\ C_2\ C_3$ $T\ H\ T$	(7) $C_1\ C_2\ C_3$ $T\ T\ H$	(8) $C_1\ C_2\ C_3$ $T\ T\ T$

Expressed as ratios, the probability of three heads is 1/8: of two heads and one tail 3/8; of one head and two tails 3/8; of three tails 1/8. The sum of these probability ratios is

$$1/8+3/8+3/8+1/8=1.00$$

The expected appearance of heads and tails can be expressed as:

$$(H+T)^3=H^3+3H^2T+3HT^2+T^3$$

in which

$1\ H^3$ =three heads ; 1 out of 8: probability ratio=1/8
$3H^2T^1$=two heads 1 tail : 3 out of 8; probability ratio =3/8
$3HT^2$=one head and two tails; 3 out of 8; probability ratio=3/8
$1\ \ T^3$=three heads : 1 out of 8; probability Ratio =1/8

Total=8 Total=1

By exactly the same method used above for two and for three coins, the expected appearance of heads and tails for four coins tossed at the same time can be put as:

$$(H+T)^4=H^4+4H^3T+6H^2T^2+4HT^3+T^4$$

in which

$1H^4$ =4 heads : 1 out of 16; probability ratio =1/16
$4H^3T$ =3 heads 1 tail : 4 out of 16; probability ratio =4/16
$6H^2T^2$=2 heads 2 tails : 6 out of 16; probability ratio=6/16
$4HT^3$ =1 head 3 tails : 4 out of 16; probability ratio =4/16
$1T^4$ =4 tails : 1 out of 16; probability ratio =1/16

Total=16 Total=1

Similarly we can determine the probability of different combinations of heads and tails when we have five, six or any number of coins. These various outcomes may be obtained in a more direct way, however, than by writing down all of different combinations which may occur. If the coin is tossed 10 times, for example, the probabilities of various combinations of heads and tails will be expressed by expansion of the binomial $(H+T)^{10}$. When the binomial $(H+T)^{10}$ is expanded, the terms are

$$H^{10}+10H^9T^1+45H^8T^2+120H^7T^3+210H^6T^4+252H^6T^5+$$
$$210H^4T^6+120H^3T^7+45H^2T^8+10H^1T^9+T^{10}$$

in which

$1H^{10}$=10 heads: 1 out of 1024; probability ratio $= \frac{1}{1024}$

$10\ H^9T^1$=9 heads 1 tail : 10 out of 1024; probability ratio $= \frac{10}{1024}$

$45H^8T^2$=heads 2 tail : 45 out of 1024; probability ratio $= \frac{45}{1024}$

$120H^7T^3$=7 heads 3 tails: 120 out of 1024; prabability ratio $= \frac{120}{1024}$

$210H^6T^4$=6 heads 4 tails: 210 out of 1024; probability ratio $= \frac{210}{1024}$

$252H^5T^5$=5 head 5 tails : 252 out of 1024; probability ratio $= \frac{252}{1024}$

$210H^4T^6$=4 heads 6 tails : 210 out of 1024; probability ratio $= \frac{210}{1024}$

$120H^3T^7$= 3 heads 7 tails : 120 out of 1024; probabilility ratio $= \frac{120}{1024}$

$45H^2T^8$=2 head 9 tails : 45 out of 1024; probability ratio $= \frac{45}{1024}$

$10H^1T^9$=1 head 9 tails : 10 out of 1024; probability ratio $= \frac{10}{1024}$

$1\ T^{10}$= 10 tails : 1 out of 1024 ; probability ratio $= \frac{1}{1024}$

Total=1024 Total=1

These data are represented graphically in figure 7 6 by a histogram and frequency polygon plotted on the same axes. The eleven terms of the expansion have been laid off at equal distances along the horizontal axes, and changes of the occurrence of each combination of heads and tails are plotted are frequencies on vertical axis.

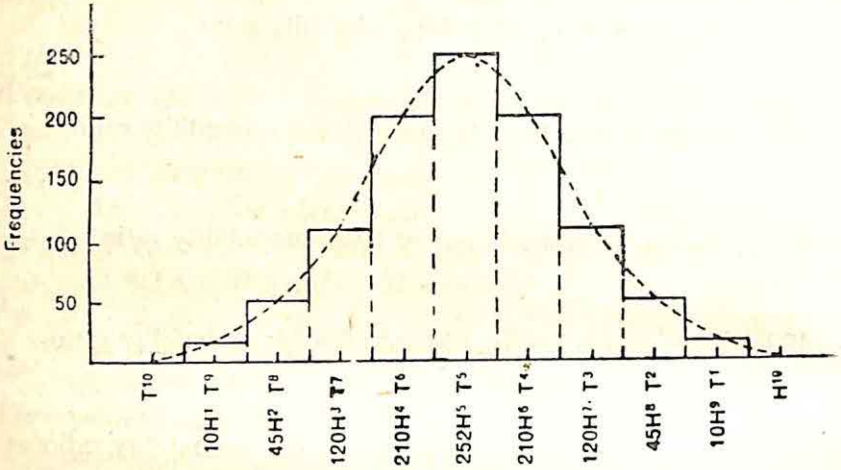


FIG. 7.6 Frequency Polygon obtained from the Expansion of $(H+T)^{10}$

The frequency curve plotted in Figure 7.6 from the expansion of the expression $(H+T)^{10}$ is a symmetrical many-sided polygon. If he number of factors (*e.g.*, coins in the present case) determining this polygon are increased from 10 to 15, 15 to 25, say, the lines which constitute the sides of polygon would increase regularly in number, becoming progressively shorter. With each increase in the number of factors the point on the frequency polygon would move together. Finally, when the number of factors become very large, the polygon would exhibit a perfectly smooth surface like that of the curve in Figure 7.7.

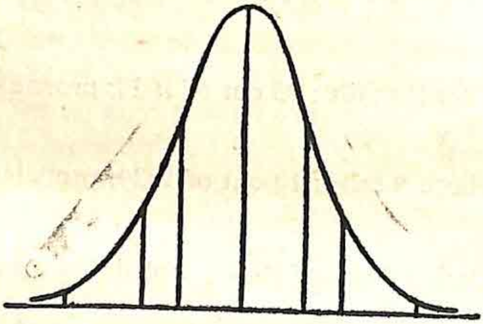
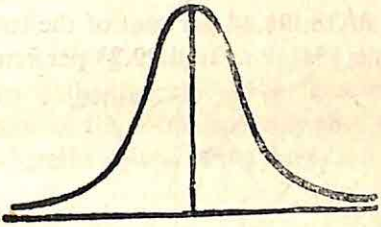


FIG. 7.7 Normal Probability Curve.

This bell-shaped curve is called the 'normal probability curve', or simply the 'normal curve', and is of great value in educational research when we make use of mental measurement. It may be noted that normal distribution is not an actual distribution of scores on any test of ability or academic achievement, but is instead, a mathematical model. The distributions of test scores approach the theoretical normal distribution as a limit, but the fit is rarely ideal and perfect.

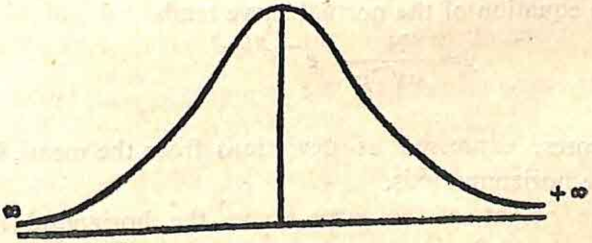
The normal probability curve has the following characteristics:

1. The curve is symmetrical around its vertical axis called ordinate. the characteristic of symmetry about the ordinate at the central point of curve implies that the size, shape and slope of the curve on one side of the curve is identical to that on the other side of the curve. The values of mean, mode and median computed for a distribution following this curve always coincide and have the same value.

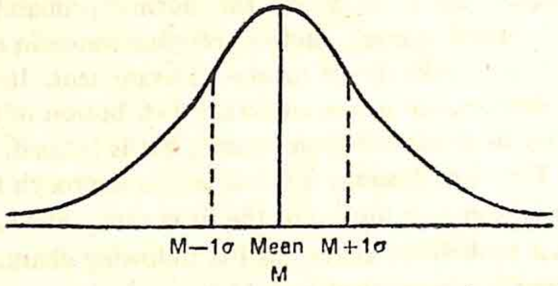


2. The height of the vertical line called ordinate is maximum at the mean and in the unit normal curve is equal to 0.3989.

3. The curve asymptotic. It approaches but does not meet horizontal axis and extends from —∝ (minus infinity) to +∝ (plus infinity).

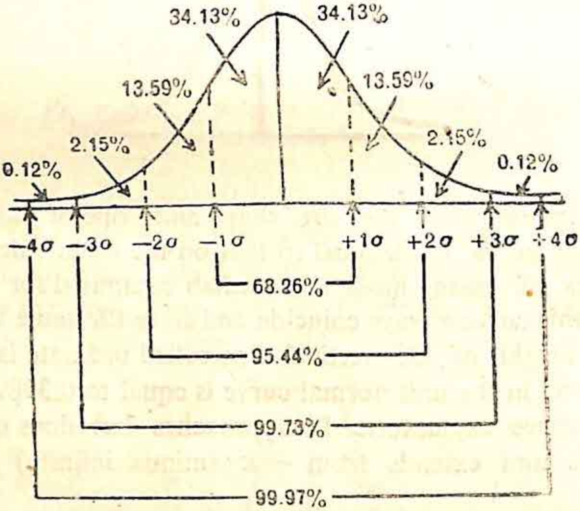


4. The points of inflection of the curve occur at points ±1 standard deviation (± 1 σ) unit above and below the mean. Thus the



curve changes from convex to concave in relation to the horizontal axis at these points.

5. 68.26 per cent of the total area of the curve falls between the limits $M+1\sigma$ and M-1σ; 95.44 per cent of the total area of the curve falls between limits $M\pm2\sigma$; and 99.73 per cent of the total area of



the curve falls between $M\pm3\sigma$.

6. The equation of the normal curve reads

$$y=\frac{N}{\sigma\sqrt{2\pi}}\ e^{-x^2/2\sigma^2}$$

in which

x=scores, expressed as deviations from the mean, laid off along the horizontal axis.

y=the height of the curve above the horizontal axis *i.e.*, the frequency of a given x value

N=number of cases in the sample of group

σ =standard deviation of the distribution

π =3.1416

$e=2.7183$

When N and σ are known, it is possible from equation of the normal curve to compute:

(*i*) the frequency (or y) of a given value x; and

(*ii*) the number, or percentage between two points, or above or below a given point in the distribution.

However these calculations are rarely necessary, as tables are available from which this information may be readily obtained. Since normal distribution plays a central role in statistical theory as well as in its application, a knowledge of the normal table (Appendix Table A) is indispensable in the solution of a number of problems. For this reason it is very essential that the use of Table *A* be clearly understood.

Table A gives the fractional parts of the total area under the normal curve found between the mean and ordinates (y's) erected at various distances from the mean. The total area under the curve is taken arbitrarily to be 10,000, because of the greater convenience with which fractional parts of the total area may then be calculated. We know that the $x(=X-M)$ measures the deviation of a raw score (X) from the mean (M). If x is divided by σ, then this deviation is expressed in σ units. These σ deviation scores are called sigma-scores or Z-scores$\left(\text{i.e } Z=\frac{X-M}{\sigma}=\frac{x}{\sigma}\right)$. The first column of the table, $\frac{x}{\sigma}$, gives distances from the mean in tenths of σ and distances from the mean in hundredth of σ are given by headings of other columns.

How to Consult Normal Table?

To find the number of cases in the normal distribution between the mean and the ordinate erected at a distance of 1 σ from the mean, we go down the $\frac{x}{\sigma}$ column until 1.0 is reached, and in the next column under .00 we take the entry opposite 1.0, namely 3413. This figure means that 3413 cases in 10,000, or 34.13 per cent of the entire area of the curve lies between the mean and 1 σ. Similarly if we have to find the percentage of the distribution between the mean and 1.45 σ, say, we go down the $\frac{x}{\sigma}$ column to 1.4, then across horizontally to the column headed .05, and take the entry 4265. This means that in a normal distribution, 42.65 per cent of the total area lies between the mean and 1.45 σ.

We have so far considered only σ distances measured in the positive direction from the mean. For this we have taken account only of the right half of the normal curve. Since the curve is symmetrical about the mean, the entries in Table A apply to distances measured in the negative direction (to the left) as well as to those measured in the positive direction. If we have to find the percentage of distribution between the mean and—1.37 a, for instance, we take entry 4147 in the column headed .07, opposite 1.3 in the $\frac{x}{a}$ clumn. This entry means that 41.47 per cent of the cases in the normal distribution fall between the mean and —1.37 a. The percentage of the cases between the mean and—0.56 a is 21.23 and so on.

For practical purposes we take the curve to end at points—3a and +3a distant from the mean as the normal curve does not actually meet the base line. Table A shows that 4986.5 cases in the total 10,000 fall between the mean and +3a; and the same number of 4986.5 cases fall between the mean and —3a. Hence, 9973 cases in 10,000 or 99.73 per cent of the entire distribution, lie within the limits —3a and +3a. The rest 0.27 per cent of the distribution beyond ± 3a is not considered except in very large samples.

Application of the Normal Curve

There are a number of applications of the normal curve in the field of educational research.

1. *To convert raw scores into standard scores.* Suppose the student *A* gets a score 45 on an arithmetic test which was administered to a group of 200 students. The mean and standard deviation of the test scores for the group are 30 and 5 respectively. Thus the standard score or sigma score of the student *A* is $Z = \frac{X-M}{a} = \frac{x}{a} = \frac{45-30}{5} = \frac{15}{5} = 3$, indicating that 45 is 3 standard deviation above the mean.

2. *To calculate the percentile rank of scores.* When scores in a distribution are assumed to be normaley distributed, percentile rank can be calculated from a sigma score. The percentile rank is defined as the percentage of scores below a given score. Suppose that the raw score of a student has been converted to a sigma score of +2.00 if his raw score was 90 in a normal distribution with mean 80 and standard deviation 5$\left(i.e.\ Z = \frac{90-80}{5} = +2.00\right)$as in the Figure 7.8 on page 229.

The percentile rank of the student is 98, for 98 per cent of the scores fall below+2.00 Z. Similarly, the sigma score of a student

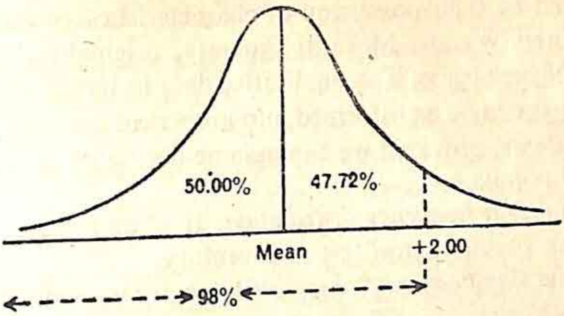


FIG. 7.8. Percentile Rank of a Student having Raw Score 90 in a Normal Distribution with Mean 80 and Standard Deviation 5.

having a raw score of 75 (say) in the same distribution mean 80 and standard deviation 5 is −1.00 $\left(i.e.\ Z= \frac{75-80}{5} = -1.00 \right)$ as in the Figure 7.9.

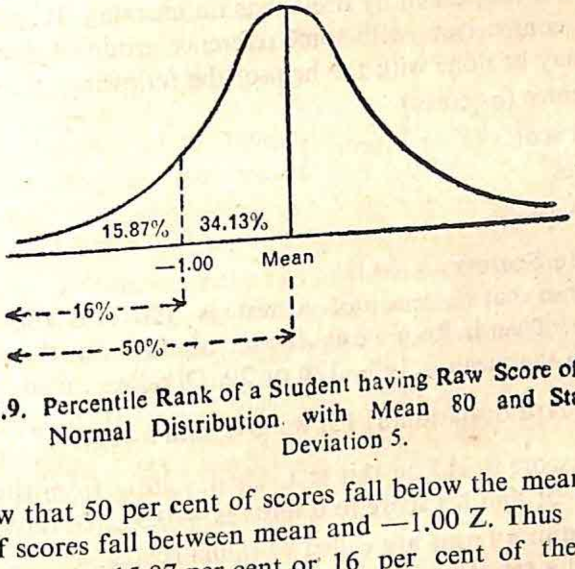


FIG. 7.9. Percentile Rank of a Student having Raw Score of 75 in a Normal Distribution with Mean 80 and Standard Deviation 5.

We know that 50 per cent of scores fall below the mean and 34.13 per cent of scores fall between mean and −1.00 Z. Thus (50 per cent −34.13 per cent)=15.87 per cent or 16 per cent of the scores fall below −1.00 Z. Hence, the percentile rank of the student is 16.

3. *To scale responses to opinionnaires, judgements, ratings or rankings by transforming them into numerical values.* We assume that the opinion or attitude, or personality trait expressed in answering a given proposition is normally distributed. From the percentage who accept each of the possible answers to a question or statement, we find a σ equivalent, which expresses the value or weight to be given to that answer. Moreover, in many psychological problems individuals are

judged or rated by their possession of characteristics or attributes not readily measured by tests. Morality, honesty, originality, etc. are the illustrations of such traits. The qualitative data in terms of such judgements or ratings can be transformed into numerical scores and scaled by finding equivalents, provided we can assume normality of distribution in the general population.

4. *To normalize a frequency distribution.* It is an important step in standardizing a psychological test or inventory.

5. *To test the significance of observations in experiments, finding their relationships with the chance fluctuations or errors that are the result of sampling procedures.*

6. *To generalize about populations from which the samples are drawn by calculating the standard error of mean and other statistics.*

MEASURES OF RELATIVE POSITIONS

A raw score on a test, taken by itself, has no meaning. It gets meaning only by comparison with some reference group of groups. The comparison may be done with the help of the following measures:

1. Sigma scores (σ scores)
2. Standard scores (Z or T scores)
3. Percentiles
4. Percentile ranks.

Sigma Scores (σ Scores)

Let us suppose that the mean of a test is 120 and the standard deviation is 15. Then if Ram earns a score of 140 on this test, his deviation from the mean is 140—120 or 20. Dividing this deviation of 20 by the standard deviation(σ) 15, we give him a score of $\frac{20}{15}$ or 1.33. If Anil's score is 115 on this test, his deviation from the mean is 115—120 or —5; and his score in σ units is -0.33. Deviations from mean expressed in σ terms are called as sigma scores.

The mean of a set of σ scores is always 0 (the reference point) and the standard deviation (σ) is always unity or 1.00. As approximately half of the score in a distribution will lie below and half above the mean, about half of our sigma scores will be negative and half positive. A sigma score makes possible a realistic comparison of scores and provides a basis for equal weighting of the scores as the scores on different tests are expressed on a scale with a mean of zero and a standard deviation of 1. The sigma scores are also useful in hypothesis testing, determination of percentile ranks, and probability judgements.

Standard Scores (*Z* and *T* Scores)

Sigma scores are often small decimal fractions and half of them are negative; and hence somewhat inconvenient to deal with in computation. For these reasons, σ scores are usually converted into a new distribution with mean and standard deviation so selected as to make all scores positive and relatively easy to handle in computation. Such scores are called standard scores and are expressed as *Z* or *T* scores.

The conversion of raw scores into standard scores is done with the help of a linear transformation which does not change the shape of the distribution in any way. The formula for the conversion of a raw score to standard score is as follows:

Let X=a scores in the original distribution
X'=a standard score in the new distribution
M and M'=mean of the raw and standard scores.
σ and σ'=SD's of raw and standard scores.

Than $\frac{X'-M'}{\sigma^1}=\frac{X-M}{\sigma}$

or $X'=\frac{\sigma'}{\sigma}(X-M)=M'$ (7.10)

When the mean (M') and standard deviation (σ') are taken to be 50 and 10 respectively, the standard score is called as *T* score.

i.e. $T=\frac{10}{\sigma}(X-M)+50$ (7.11)

An example will illustrate the use of these formulae. Given a distribution with mean=86, σ=15. Anil's score is 90 and Jamil's 80. Express these raw scores as (1) standard scores in a distribution with mean of 500 and σ of 100 and (*2*) *T* score.

By formula (7.10)

$X'=\frac{100}{15}(X-86)+500$

Substituting Jamil's score of 80 we have

$X'=6.67\ (80-86)+500=460.$

Substituting Anil's score of 90 we have

$X'=6.67\ (90-86)+500=527.$

By formula (7.11)

$T=\frac{10}{15}(X-86)+50$

Substituting Anil's score of 90, we have

$T=.66\ (90-86)+50=53$

Substituting Jamil's score of 80, we have

$T=.66\ (80-86)+50=46$

Percentiles

Percentiles are the points which divide the entire scale of measurement into 100 equal parts. They are denoted by P_0, P_1, P_2, P_3, P_4,..., P_{50} .....,P_{99}, P_{100}.

First percentile may be defined as that point in a frequency distribution below which lie 1 per cent of the total measures of scores. Similarly, tenth percentile may be defined as that point in a frequency distribution below which 10 per cent of the total measures or scores fall. It is evident that the median, expressed as a percentile, is P_{50}. It should be noted that P_0 lies at the beginning of the distribution and P_{100} at the end of the distribution.

The method of calculating percentiles is essentially the same as that used in finding the median. The formula is

$$P_p = l + \left(\frac{pN - F}{f_p} \right) \times i \tag{7.12}$$

in which

P_p=percentage of the distribution wanted, *e.g*, 5 per cent, 10 per cent etc

l=exact lower limit of the class interval upon which P_p lies

$_p$N= part of N to be counted off in order to reach P_p

F=sum of all scores upon intervals below l

f_p=number of scores within the interval upon which P_p falls

i=lenght of the class interval.

The following example will illustrate the use of formula (7.12). Calculate P_{30}, P_{60}, and P_{75} for the following distribution.

TABLE 7.12. The calculation of Percentiles from Data Grouped into a Frequency Distribution

Scores	*(f)*	*Cumulative frequency (F)*
129.5-134.5	2	40
124.5-129.5	0	38
119.5-124.5	1	38
114.5-119.5	3	37
109.5-114.5	5	34
104.5-109.5	7	29
99.3-104.5	6	22
94.5- 99.5	2	16
89.5- 94.5	7	14
84.5- 89.5	5	7
79.5- 84.5	2	2
	N=40	

For computing P_{30} we have to first find $_{p}N$

Here, 30 per cent of 40 is 12

New l=89.5 is the lower limit of the class interval in which P_{30} (=12) lies.

$F=7;\ f_p=7$ and $i=5$

Using formula (7.12)

$$P_{30}=89.5+\left(\frac{12-7}{7}\right)\times 5$$

$$=89.5+3.57=93.07$$

Similarly $$P_{60}=104.5+\left(\frac{24-22}{7}\right)\times 5$$

$$=104.5+1.43=105.93$$

$$P_{75}=109.5+\left(\frac{30-29}{5}\right)\times 5=109.50+1.00=110.50$$

Percentile Ranks

The percentile rank is the point in the distribution below which a given percentage of scores fall. If the 90th percentile rank is a score of 70, 90 per cent of the scores fall below 70. The median is the 50th percentile rank, for 50 per cent of the scores fall below it.

In general, we may say that if K per cent of the members of a sample have scores less than a particular point value X, then X is the Kth percentile and K is the percentile rank of X.

The calculation of the percentile ranks is the reverse process of calculation of percentile to points. We have to calculate ranks corresponding to particular scores. In case of individual scores, first we find the rank of individual scores from the bottom *i.e.* rank in ascending order. If R is this rank and N is the total number of cases then,

$$\text{Percentile Rank}=100-\frac{100R-50}{N} \tag{7.13}$$

Anjni ranks 7th in her class of 140 students. Six students rank above her, 133 below her. Her percentile rank is:

$$100-\left(\frac{100\times 7-50}{140}\right)=100-4.64$$

$$=100-5$$

$$=95$$

MEASURES OF RELATIONSHIP

We have so far considered the statistical description of a single variable. Now we shall study the problem of describing the degree of simultaneous variation of two variables. The data in which we secure measures of two variables for each individual is called a bivariate data. For example, we get a bivariate data if we have measures of both height and weight for a group of school children. The essential feature of the bivariate data is that one measure can be paired with another measure for each member of the group. When we study bivariate data we may like to know the degree of relationship between variables of such data. This degree of relationship is known as correlation. It can be represented quantitatively by the coefficient of correlation.

We observe that students who have high intelligence quotients tend to receive high scores on an achievement test in arithmetic, whereas those with low intelligence quotients tend to score low. When this type of relationship is observed, the variables of intelligence and attainment in arithmetic are said to be positively correlated. In general, when the individuals above average in one variable tend also to be well above average in the other; those below average in the one tend to be correspondingly below average in the other; and those at or near the average in one tend also to be at or near the average in the other; the two variables (or measures) show a positive correlation

Sometimes students making high scores in one variable are likely to make low scores in another variable and vice versa. To illustrate, suppose that in a class of 10 students, the student who stood first in an achievement test of mathematics was ranked lowest on a test of anxiety, the student who stood second in mathematics test ranked next to the bottom (ninth) in test of anxiety; and that each student stood just as far from the top in mathematics test as form the bottom in anxiety test. Here the correspondence between ranks in mathematics test and anxiety test is regular and definite enough, but the direction of relationship is inverse (negative).

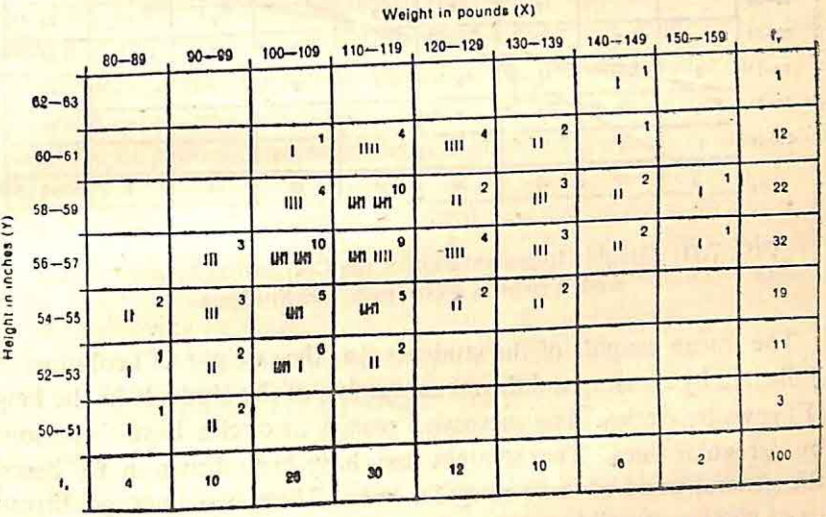
When the relationship between two sets of variables is a pure chance relationship, we say there is no correlation. For example, in a class of 10 students a student with a high score in English test is likely to be anywhere within the total range in terms of his score in a test of numbers. The three students scoring first three positions in the English test may score fifth, seventh and fourth positions on the number test. The three lowest students in English test may score third, eighth and sixth positions in the number test.

Coefficient Correlation

The intensity or degree of linear correlation is represented quantitatively by coefficient of correlation. Its value ranges from — 1.00 to +1.00. A value of —1.00 describes a perfect negative correlation and +1.00 describes perfect positive correlation. A zero value describes complete lack of correlation between the two variables. The sign of the coefficient indicates the direction of the relationship and the numerical value its strength.

The Scattergram and the Correlation Table

When we have a large data, much time and energy will be saved by arranging the data in both variables X and Y in the form of a diagram or chart which is called a Scattergram or a Scatterdiagram. It represents a bivariate distribution by taking one variable X along horizontal line and the other variable Y along the vertical line. Suppose that we have records for 100 students, giving the heights (X) and weights (Y) of students; then to each student corresponds a pair of values (X_i, Y_i).



Weight in pounds (X)

Height in inches (Y)	80—89	90—99	100—109	110—119	120—129	130—139	140—149	150—159	f_y
62—63							I 1		1
60—61			I 1	IIII 4	IIII 4	II 2	I 1		12
58—59			IIII 4	卌 卌 10	II 2	III 3	II 2	I 1	22
56—57		III 3	卌 卌 10	卌 IIII 9	IIII 4	III 3	II 2	I 1	32
54—55	II 2	III 3	卌 5	卌 5	II 2	II 2			19
52—53	I 1	II 2	卌 I 6	II 2					11
50—51	I 1	II 2							3
f_x	4	10	26	30	12	10	6	2	100

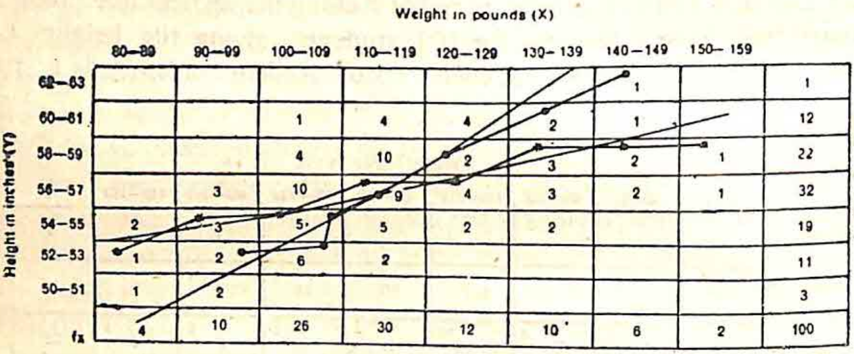
FIG. 7.10. A Scattergram Showing Paired Heights and Weights of 100 Students.

Each of the 100 students can be represented on the diagram with respect to height and weight. Suppose that a student weights 120 pounds and is 60 inches tall. His weight locates him in the fifth column from the top. Accordingly, a'tally' is placed in the second cell of the fifth column. There are four tallies in all in this cell, that is, there are four students who weigh from 120-129 pounds, and are 60-51 inches tall. Each of the 100 students is represented by a tally in a cell of the table in accordance with two characteristics of height and weight.

Along the bottom of the diagram in the f_x row is tabulated the number of students who fall in each weight interval; while along the righthand margin in the f_y column is tabulated the number of students who fall in each height interval. The f_y column and f_x row must each total 100, the total number of students. After listing all the tallies, the frequency in each cell is added and entered on the diagram. The scattergram is also called a correlation table.

The Regression Lines

The scattergram in Figure 7.11 represents the linear relationship between height and weight in a group of 100 college male students graphically.



	80–89	90–99	100–109	110–119	120–129	130–139	140–149	150–159	
62–63							1		1
60–61			1	4	4	2	1		12
58–59			4	10	2	3	2	1	22
56–57		3	10	9	4	3	2	1	32
54–55	2	3	5	5	2	2			19
52–53	1	2	6	2					11
50–51	1	2							3
f_x	4	10	26	30	12	10	6	2	100

FIG. 7.11. Graphic Representation of the Correlation between Height and Weight in a Group of 100 Students.

The mean heights of the students in the weight (X) columns are indicated by crosses, and the mean weights of the students in the height (Y) rows by circles. The successive crosses or circles have been joined by irregular lines. Two straight lines have been drawn in to describe the general trend of these irregular lines. These two lines go through, or as close as possible to, the crosses or the circles. Such lines are called the regression lines or the lines of best fit. The line drawn to represent the trend of crosses in Figure 7.11 is the line of regression of Y on X, and it is the line from which the sum of the squares of deviations of all coordinate points (crosses) has the smallest value. The other line drawn to represent the trend of circles is the line of regression of X on Y, and it is the line from which the sum of the squares of deviations of all coordinate points (circles) is the minimum.

As the coefficient of correlation approaches zero (0), the coordinate points fall farther from the regression lines. When it is + 1.00, all of

Y on X, and it is the line from which the sum of the squares of every increase in X there is a proportional increase in Y; the value of —1.00 indicates that for every increase in X there is a proportional decrease in Y. The greater the angle separating the two lines, the lesser is the correlation between two variables X and Y.

Methods of Correlation

There are various methods of correlation. Their use is relative to the situation and type of data. We may have data in scores. There are many situations in which the researcher does not have scores and has to work with data in which differences in a given attribute can be expressed only by ranks; or by classifying an individual into one of several descriptive categories. Some of these methods of correlation are described as under:

Product-moment correlation In some situations the data for two variables X and Y are expressed in interval or ratio level of measurement and the distributions of these variables have a linear relationship. Moreover, the distributions of the variables are uni-modal and their variances are approximately equal. In such situations we may make use of product moment method of correlation. It is also called Pearson's r

1. *Calculation of Pearsons 'r' from ungrouped data.* When the N (size of the sample or group) is small or the raw scores are small numbers, there is no need of grouping the data and we may make use of the following formula:

$$r_{XY}=\frac{N\,\Sigma xy-\Sigma x\,.\,\Sigma y}{\sqrt{[N\Sigma x^2-(\Sigma x)^2]\,[N\Sigma y^2-(\Sigma y)^2]}} \qquad (7.14)$$

in which

x=Deviation of the X measures from the assumed mean.

y=Deviation of the Y measures from the assumed mean.

To illustrate the use of this formula, let us compute product-moment r from the following data in the two variables X and Y for 11 students.

X : 45 55 56 58 60 65 68 70 75 80 85
Y : 56 50 48 60 62 64 65 70 74 82 90

Using formula (7.14) for the data in the Table 7.13:

$$r_{XY}=\frac{11\times1403-2\times6}{\sqrt{[11\times1414-(2)^2]\,[11\times1650-(6)^2]}}$$

$$=0.92$$

TABLE 7.13. The Calculation of Product Moment Correlation from Ungrouped Data when Deviations are taken from Assumed Mean

X	Y	x	y	x^2	y^2	xy
45	56	-20	-9	400	81	180
55	50	-10	-15	100	225	150
56	48	-9	-17	81	289	153
58	60	-7	-5	49	25	35
60	62	-5	-3	25	9	15
65(AM)	64	0	-1	0	1	0
68	65(AM)	3	0	9	0	0
70	70	5	5	25	25	25
75	74	10	9	100	81	90
80	82	15	17	225	289	255
85	90	20	25	400	625	500
		$\Sigma x=2$	$\Sigma y=6$	$\Sigma x^2=1414$	$\Sigma y^2=1650$	$\Sigma xy=1403$

when a calculating machine is available, the value of Pearson's r can be obtained from the original measures with the help of the following formula:

$$r_{XY}=\frac{N\,\Sigma XY-\Sigma X.\,\Sigma Y}{\sqrt{[N\,\Sigma X^2-(\Sigma X)^2]\,[N\,\Sigma Y^2-(\Sigma Y)^2]}} \qquad (7.15)$$

2. *Calculation of Pearson's 'r' from grouped data.* When N is large or even moderate in size, and when no calculating machine is available, the best procedure is to group data in both variables X and Y and to from a scattergram. The values from the scattergram may be used in the following formula:

$$r_{XY}=\frac{N\,\Sigma fxy-\Sigma fx.\,\Sigma fy}{\sqrt{N\,\Sigma fx^2-(\Sigma fx)^2]\,[N\,\Sigma fy^2-(\Sigma fy)^2]}} \qquad (7.16)$$

To illustrate the use of formula (7.16) consider the scattergram given on page 239.

The computation for the values of Σfx, Σfx^2, Σfxy, etc. may be done by following steps as under:

Step 1. The distribution of general science test scores for the 85 students is found in the f_y column at the right of the scattergram. Assume a mean for the distribution of scores of general science (the mid-point of that interval which contains the largest frequency), and draw double lines to mark off the row in which the assumed mean falls. In the present example, the mean for the general science scores

Arithmetic (X)

General Science (Y)

	12—13	14—15	16—17	18—19	20—21	22—23	24—25	f_y	y	fy	fy^2	fxy
35—37					1 (0) 0		1 (6) 6	2	+3	6	18	6
32—34					6 (0) 0	3 (2) 6		9	+2	18	36	6
29—31		1 (−3) −3	2 (−2) −4	6 (−1) −6	8 (0) 0	1 (1) 1		18	+1	18	18	−12
26—28		4 (0) 0	4 (0) 0	6 0 0	11 (0) 0	4 (0) 0	1 (0) 0	30	0	0	0	0
23—25	2 (4) 8	1 (3) 3	6 (2) 12	5 (1) 5	4 (0) 0	1 (−1) −1		19	−1	-19	19	27
20—22	3 (8) 24	2 (6) 12	1 (4) 4	1 (2) 2				7	−2	−14	28	42
f_x	5	8	13	18	30	9	2	85		9	119	69
x	−4	−3	−2	−1	0	+1	+2					
fx	−20	−24	−26	−18	0	9	4	-75				
fx^2	80	72	52	18	0	9	8	239				
fxy	32	12	12	1	0	6	6	69				

FIG. 7.12. A Scattergram Showing Paired Scores of 85 Students on the Tests of Arithmetic and General Science.

has been taken at 27 (mid-point of interval 26-28) and y's (deviations from the assumed mean) have been taken from this point. Now fill in the fy and fy^2 columns.

Step 2. The distribution of the arithmetic test scores of 85 students is found in the f_x row at the bottom of the scattergram. Assume a mean for this distribution, and draw double lines to designate the column under the assumed mean. The mean for the arithmetic scores is taken at 20.5 (mid-point of interval 20-21), and the x's (deviations from the assumed mean) are taken from this point. Fill in the fx and fx^2 rows.

Step 3. The fxy for a cell is computed by multiplying the frequency given in the particular cell with the corresponding x and y. For example, there is a frequency 1 in the cell corresponding to arithmetic score interval 24-25 and general science score interval 35-37. The corresponding x for this cell frequency is + 2 and the corresponding y is + 3. Thus fxy for this cell is $1 \times 2 \times 3 = 6$. Similarly the value for fxy is computed for all the cells and their sum Σfxy is calculated row-wise as well as column-wise. The two sums should equal each other. In the present example it has come equal to 69.

Step 4. Substituting the values for Σfx; Σfx^2; Σfy, Σfy^2 and Σfxy in the formula (7.16), we get:

$$r_{xy}=\frac{(85)(69)-(-75)(9)}{\sqrt{[(85)(239)-(-75)^2][(85)(119)-(9)^2]}}$$
$$=0.54$$

Rank order correlation. Rank order correlation is a simplest method of correlation. It is also known as the Spearman rank order coefficient of correlation and is denoted by ρ (rho). This method is used when the data are available only in ordinal form of measurement (ranked) rather than in interval or ratio form, or if the number of paired variables is more than 9 or fewer than 30.

To compute the Spearman rank order coefficient of correlation, the following formula is used:

$$\rho=1-\frac{6\ \Sigma\ D^2}{N(N^2-1)} \tag{7.17}$$

in which

D=the difference between paired ranks

D^2=the sum of the squared differences between ranks

N=number of paired ranks.

To illustrate the calculation of rank difference correlation coefficient, two judges *X* and *Y* ranked 10 students in a declamation contest. The ranks given by the two judges ara given in the Table 7.14.

TABLE 7.14. The Calculation of Rank Difference Correlation

Students	*Rank assigned by X*	*Rank assigned by Y*	*D*	*D*
A	1	2	−1	1
B	5	3	2	4
C	4	5	−1	1
D	6	4	2	4
E	3	6	−3	9
F	2	1	1	1
G	7	7	0	0
H	9	8	1	1
I	10	9	1	1
J	8	10	−2	4
				$\Sigma D^2=26$

Using formula (7.17)

$$\rho=1-\frac{6\times 26}{10(100-1)}$$
$$=0.84$$

Biserial and point biserial correlations. In many problems we may wish to compute the correlation between traits and other attributes when the members of the group can be measured in one variable, but can be classified into only two categories in the second or dichotomous variable. The term dichotomous means 'cut into two parts'. A researcher, for example, wishes to know the correlation between general intelligence and social adjustment in a group of primary school children. The children can be measured in the first variable (in terms of scores on a test of intelligence), but they can be simply classified as socially adjusted or socially maladjusted on the second variable. In such a situation the researcher may compute a 'biserial correlation' between the set of scores on the intelligence test and two category groupings—socially adjusted and socially maladjusted. Other instances of two-fold classification with reference to some attribute are athletic-nonathletic, drop outs and stay ins, and so on.

When the variable which has been classified into two categories can be thought of having a genuine dichotomy, we make use of 'Point biserial correlation'. Examples of genuine dichotomies are male-female, living-dead, psychotic-normal, colour blind-normal and so on.

Biserial and point biserial correlations are both useful in item analysis in test construction.

The formula for computing the biserial coefficient of correlation (r_{bis}) is

$$r_{bis}= \frac{M_p - M_q}{\sigma t} \times \frac{pq}{y} \tag{7.18}$$

in which

M_p=mean of the group in the 1st category usually the group showing superior or desirable characteristics.

M_q=mean of the group in the second category.

σ_t=standard deviation of the entire group.

p=proportion of the entire group in category 1.

q=proportion of the entire group in category 2.

y=height of the normal curve ordinate dividing the two parts, p and q.

To illustrate the use of the formula (7.18), let us compute the r_{bis} for the distribution of scores on an achievement test in the Table 7.15 earned by those students who answered 50 per cent or more and those who answered less than 50 per cent of the items in an achievement test in arithmetic correctly.

M_p, the mean of group of students answering 50 per cent or more of the items correctly (N_1=120) has come out to 156.75 and M_q, the

TABLE 7.15. The Calculation of Biserial Correlation for the Data Grouped into a Frequency Distribution

Achievement test scores	*Students answering 50 per cent or more of the items correctly (Group I)*	*Students answering less than 50 per cent of the items correctly (Group II)*	*Total*
185 194	7	0	**7**
175-184	16	0	**16**
165-174	10	6	**16**
155-164	35	15	**50**
145-154	24	40	**64**
135-144	15	26	**41**
125-134	10	13	**23**
115-124	3	5	**8**
105-114	0	5	**5**
	$N_1=120$	$N_2=110$	$N=230$

mean of group of students answering less than 50 per cent of the items correctly ($N_2=110$) to be 144.05. The standard deviation, for the whole group ($N=230$) is 16.81.

There are 120 or 0.52 of the sample in group I ($p=120/230$); and 110 or 0.48 in group II ($q=110/230$). Assuming a normal distribution an arbitrary cut is imposed at the point 0.52 below the mean, as shown in the Figure 7.13. The height of the ordinate y, as read from the normal table is 0.3984

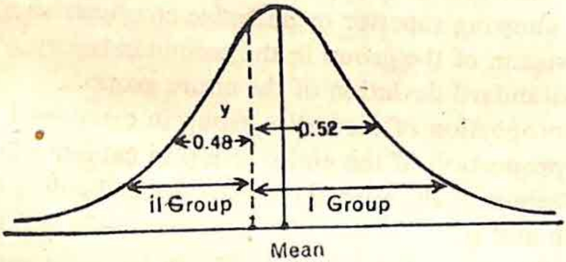


FIG. 7.13 Graphical Representation of the Areas Covered by groups I and II.

Using formula (7.18)

$$r_{bis}=\frac{156.75-144.05}{16.81}\times\frac{0.52\times0.48}{0.3984}$$

$$=0.47$$

In many situations, a more convenient formula for biserial r is

$$r_{bis} = \frac{M_p - M_t}{\sigma_t} \times \frac{p}{y} \tag{7.19}$$

in which

M_t=mean of the entire group.

M_p, σ_t, p and y have the same meaning as in formula (7.18). Using formula (7.19) in the above example, we have:

$r_{bic} = 0.47$

The formula for computing the point biserial correlation (r_{pbis}) is

$$r_{pbis} = \frac{M_p - M_q}{\sigma_t} \times \sqrt{pq} \tag{7.20}$$

in which

M_p, M_q σ_t, p and q have the same meaning as in formula (7.18)

To illustrate the use of the formula (7.20), let us compute r_{pbis} for an item of a test administered to a group of 15 students. The total test score and response of each student to the item are given in the Table 7.16. The number 1 means the item was passed and a 0 that it was failed.

TABLE 7.16. The Calculation of Point Biserial Correlation between the Total Test Score and an Item Score

Students	*Total Test Score*	*Item Score*
1	25	1
2	23	1
3	18	0
4	24	0
5	23	1
6	20	0
7	19	0
8	22	1
9	21	1
10	23	1
11	21	0
12	20	0
13	21	1
14	21	1
15	22	1
Sums	323	9

M_p, the mean of the group of students passing the item (N_1=9) is got by adding their total scores on test and then dividing the sum by

9 and M_q, the mean of the group of students failing the item ($N_2=6$).

$$M_p = \frac{201}{9} = 22.33$$

$$M_q = \frac{122}{6} = 20.33$$

$$M_t = \frac{323}{15} = 21.53$$

$$\sigma_t = 1.82$$

$$p = \frac{9}{15} = 0.60$$

$$\text{and } q = \frac{6}{15} = 0.40.$$

Assuming a normal distribution an arbitrary cut is imposed at the point 0.60 below the mean, as shown in the Figure 7.14. The value of y at the point of dichotomy as read from the Normal Table is 0.386

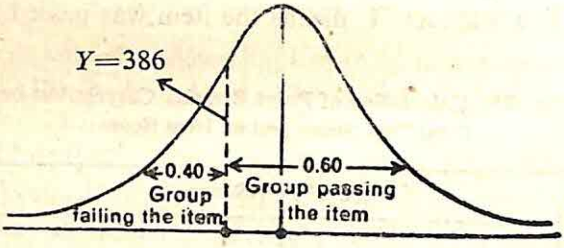


FIG. 7.14. Graphical Representation of the Area Covered by the Groups Failing and Passing an Item.

$$r_{pbic} = \frac{22.33-20.33}{1.82} \times \sqrt{0.60 \times 0.40}$$

$$=0.54$$

An alternative formula for computing the r_{pbis} is

$$r_{pbis} = \frac{M_p - M_t}{\sigma_t} \times \sqrt{\frac{p}{y}} \qquad (7.21)$$

in which

M_p, M_t, σ_t, p and y have their usual meaning.

Using formula (7.21) in the above example we get:

$$r_{pbis} = \frac{22.33-21.53}{1.82} \times \sqrt{\frac{0.60}{0.386}}$$

$$=0.54$$

Tetrachoric and phi-coefficient correlations. Tetrachoric correlation, r_t, is especially useful when a researcher wishes to find the relationship between attributes neither of which is measurable in scores. but both of which are capable of being separated artificially into two categories. For example, we may compute tetrachoric correlation between intelligence and social adjustment in students studying in the seventh grade. Students might be classified as above or below average in intelligence, on the one hand, and as socially well adjusted or socially poorly adjusted, on the other.

When the two variables to be correlated are genuinely dichatomous and the classification of variables is truly discrete (the two classes are separated by real gap between them), the phi-coefficient (ϕ-coefficient) of correlation may be used. For example, we may compute ϕ—coefficient to find out the correlation between intellectual status (feeble minded-normal) and marital status (married-unmarried). Phi-coefficient may be used also with continuous variables which have been grouped into two categories, as for instance when two tests are split at the median. It may be noted that phi-coefficient bears the same relation to tetrachoric r as the point biserial r bears to biserial r.

The formula for computing the tetrachoric r is

$$r_t = \cos\left[\frac{180^\circ \times \sqrt{BC}}{\sqrt{AD}+\sqrt{BC}}\right] \qquad (7.22)$$

in which

A, B, C and D represent frequencies in the four-fold table as shown in the Figure 7.15 below.

	X—Variable		Total
Y—Variable	B	A	A+B
	D	C	C+D
Total	B+D	A+C	

FIG. 7.15. Schematical Representation of four fold Table for the Calculation of Tetrachoric Correlation.

To illustrate the use of formula (7.22), consider the data in Table 7.17 to study the relationship between social adjustment and vocational adjustment of 100 secondary school teachers of a particular district.

TABLE 7.17. The Calculation of Tetrachoric Correlation between Vocational Adjustment and Social Adjustment

	Vocational Adjustment		
	Vocationally well adjusted	Vocationally poorly adjusted	Totals
Socially well adjusted	24 (B)	36 (A)	60 A+B
Socially poorly adjusted	28 (D)	12 (C)	40 C+D
Totals	52 (B+D)	48 (A+C)	

Using formula (7.22)

$$r_t=\cos\left(\frac{180\times{}^{\circ}\sqrt{24\times 12}}{\sqrt{36\times 28}+\sqrt{24\times 12}}\right)$$

$=\cos 62^\circ$ approx.

$=0.47$

The formula for computing the ϕ-coefficient of correlation is

$$\phi=\frac{AD-BC}{\sqrt{(A+B)(C+D)(B+D)(A+C)}} \qquad 7.23$$

in which

A, B, C and D represent frequencies in the four-fold table shown in Figure 7.15

The computatian of ϕ-coefficient from the following data collected on the basis of the responses of 100 students to the items 1 and 2 in a personality inventory illustrates the use of the formula (7.23).

TABLE 7.18. The Calculation of Phi-Coefficient Correlation between the Responses to the Items 1 and 2 of a Personality Inventory

		Item 1		
		No	Yes	Totals
Item 2	Yes	83 (B)	187 (A)	270 (A+B)
	No	102 (D)	93 (C)	195 (C+D)
Totals		(185) (B+D)	280 (A+C)	465

Using formula (7.23)

$$\phi = \frac{(102)\,(187)-(83)\,(93)}{\sqrt{(270)\,(195)\,(185)\,(280)}}$$

$$=0.22$$

The coefficient of contingency. The coefficient of contingency (C) provides a measure of correlation when each of two variables under study have been classified into two or more categories. For example, we may apply this methed to find the relationship between eye colours of fathers and eye colours of sons. The eye colours of both the groups are classified into blue, grey, hazel and brown.

The formula for computing the coefficient of contingency is:

$$C=\sqrt{\frac{S-N}{S}} \qquad (7.24)$$

in which

$S=$ the sum of quotients

$$=\Sigma\frac{(\text{observed frequency in each cell})^2}{\text{expected frequency in each cell}}$$

$N=$the size of the sample,

To illustrate the use of the formula (7.24), we may compute the coefficient of contingency to find out the relationship between nationality of the individuals and the kind of the music they prefered, from the data given in the Table 7.19.

TABLE 7.19. The Calculation of Coefficient of Contingency between the kind of Music Preferred and Nationality

		Kind of Music Preferred				
		English	*French*	*German*	*Indian*	*Totals*
Nationality of Individuals	English	(1) 60	(2) 20	(3) 10	(4) 10	100
	French	(5) 20	(6) 50	(7) 10	(8) 20	100
	Germany	(9) 10	(10) 20	(11) 10	(12) 60	100
	Indian	(13) 20	(14) 20	(15) 50	(16) 10	100
	Totals	110	110	80	100	400

Calculation of Quotient (S)

$$\text{Cell } 1 = \frac{(60)^2}{\frac{110 \times 100}{400}} = 130.91$$

$$\text{Cell } 2 = \frac{(20)^2}{\frac{110 \times 100}{400}} = 14.55$$

……………………………………

……………………………………

……………………………………

……………………………………

……………………………………

……………………………………

……………………………………

……………………………………

……………………………………

$$\text{Cell } 15 \quad \frac{(50)^2}{\frac{80 \times 100}{400}} = 125.00$$

$$\text{Cell } 16 = \frac{(10)^2}{\frac{100 \times 100}{400}} = 4.00$$

$S = 130.91 + 14.55 + \ldots\ldots\ldots + 125.00 = 4.00$

$= 606.21$

$N = 400.00$

Using formula (7.24)

$$C = \sqrt{\frac{(606.21)-(400)}{606.21}}$$

$= 0.583.$

Partial correlation. A partial correlation between two variables is one that partials out or nullifies the effects of a third variable (or a number of other variables) upon both the variables being correlated. The correlation between height and weight of male children in a group where age is permitted to very would be higher than the correlation between height and weight for a group at constant age because certāin children are older and so they are both heavier and taller. Age is a variable that enhances the magnitude of correlation between height and weight. With age held constant, the correlation would still be positive and significant but of lesser strength.

If a researcher wishes to know the correlation between height and weight with the effect of age ruled out, he could keep samples separated and compute correlation at each age level. But the partial correlation technique enables him to achieve the same result without separating data into homogenous age groups.

When only one variable is held constant, we compute first-order partial correlation. The formula is

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\ {}_{23}}{\sqrt{(1-r^2{}_{13})\,(1-r^2{}_{23})}} \qquad (7.25)$$

In a group of children from 9 to 15 years, the correlation between a general intelligence test and academic achievement (r_{12}) was found to be 0.80. Between the general intelligence test and age (r_{12}) in the same group is 0.70; and the correlation between school achievement and age (r_{23}) is 0.60.

Using the formula (7.25), the partial correlation is:

$$r_{12\cdot3}=\frac{(.80)-(.70)\quad(.60)}{\sqrt{[1-(.70)^2][1-(.60)^2]}}$$

$$=0.67$$

With the influence of age upon both general intelligence and academic achievement ruled out or nullified, then, the correlation between the two is 0.67.

With we hold two variables constant at the same time, we call the coefficient a second-order partial correlation.

Multiple correlation. So far we have been dealing with correlations between two variables at a time. At times one variable is found related with, or dependent upon, more than one variable at the same time. When we take some variables (X_2, X_3,.........) as being causes of another one (X_1) we call the one variable (X_1), the dependent variable and ones upon which it depends, the independent variables (X_2, X_3,.........). The independent variables are so called because we can manipulate, control or observe them at will and, in consequence, we expect the dependent variable to vary accordingly. The multiple correlation is related to the intercorrelations among in dependent variables as well as to their correlations with the dependent variable.

The coefficient of multiple correlation is denoted by the symbol R. It indicates the strength of relationship between one variable (independent variable) and two or more others (dependent variables) taken together. For example, we may compute multiple correlation (R) in finding the relationship of academic achievement (dependent variable) of high school students with their general intelligence and study habits (independent variables) taken together.

The formula for finding the multiple coefficient of correlation when he have three variables X_1, X_2 and X_3 is:

$$R_{1(23)}=\sqrt{1-\frac{\sigma^{2}{}_{1\cdot 23}}{\sigma_1{}^2}} \qquad (7.26)$$

in which

$1=X_1=$ Dependent variable

2 and $3=X_2$ and $X_3=$Independent variables taken together

$\sigma_1=$Standard deviation of the X_1 measures

$\sigma_{1\cdot 23}=\sigma_1\sqrt{1-r_{12}{}^2}\sqrt{1-r^2{}_{13\cdot 2}}$

To illustrate the use of the formula (7.26) we consider a problem of a researcher who wishes to study the multiple correlation of self-esteem with achievement-motivation and locus of control from the primary data given as under :

Table 7.20. Means, Standard Deviations and Intercorrelations

1. Self Esteem (X_1)	*2. Achievement Motivation (X_2)*	*3. Locus of Control (X_3)*
$X_1=119.73$	$X_2=11.26$	$X_3=64.80$
$\sigma_1=19.64$	$\sigma_2=2.89$	$\sigma_3=7.05$
$r_{12}=0.54$	$r_{13}=0{\cdot}37$	$r_{23}=0.26$

The multiple coefficient of correlation is computed in the following steps :

1. $\sigma_{1\cdot 23}=\sigma_1\sqrt{1-r^2_{12}}\ \sqrt{1-r^2{}_{13\cdot 2}}$

$$=19.64\times 0.84\times 0.96$$
$$=15.84$$

in which

$$\sigma_1=19.64 \text{ and } r_{13\cdot 2}=\frac{.37-(.54)\,(.26)}{\sqrt{1-(.54)^2}\sqrt{1-(.26)^2}}$$
$$=0.28$$

$$\sqrt{1-r^2_{12}} = \sqrt{1-(.54)^2}$$
$$=0.84$$

$$\sqrt{1-r^2{}_{13\cdot 2}} = \sqrt{1-(.28)^2}$$
$$=0.96$$

Using formula (7.26), the multiple correlation is :

2. $$R_{1(23)} = \sqrt{1-\frac{\sigma^2_{12\cdot3}}{\sigma_1{}^2}}$$

$$= \sqrt{1-\frac{(15.84)^2}{(19.64)^2}}$$

$$=0.59$$

Multiple R of 0.59 tells us to what extent self-esteem (X_1) is determined by the combined effect of achievement-motivation (X_2) and locus of control (X_3) or to what extent self esteem is related to achievement-motivation and locus of control taken together.

The Correlation Ratio

The correlation ratio is a very general index of correlation particularly adapted to data in which the relationship between the paired

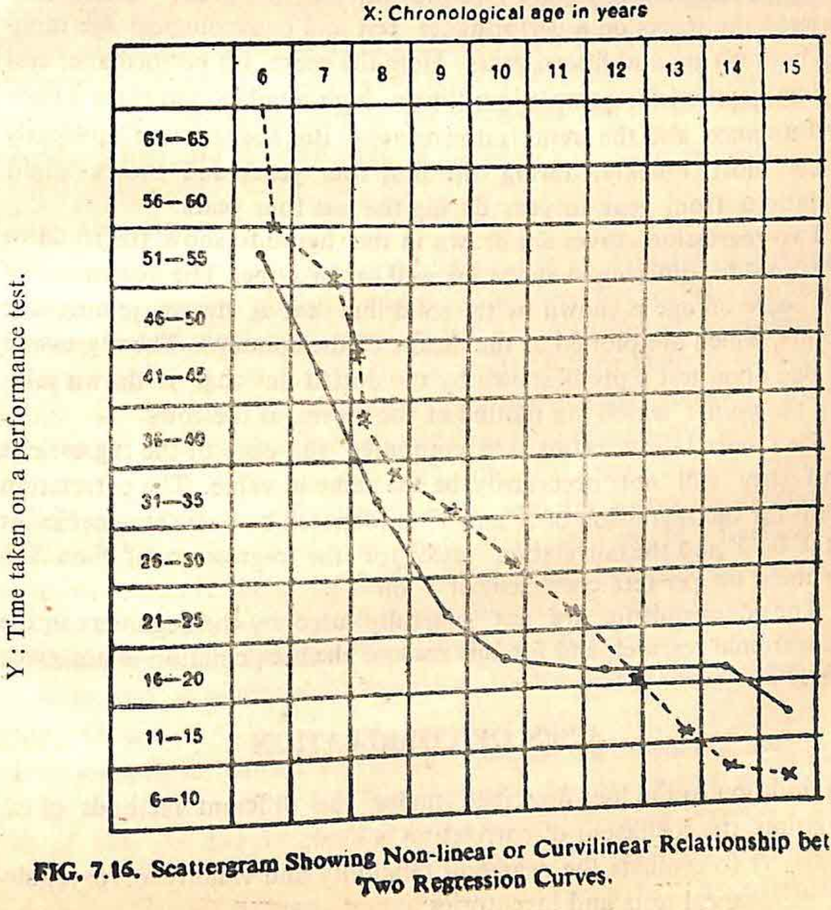


FIG. 7.16. Scattergram Showing Non-linear or Curvilinear Relationship betwee
Two Regression Curves.

values of two sets of measures, X and Y, may be described in a general way as 'non-linear'. It is the case when the drift or trend of the means of the arrays of the successive columns and rows in a correlation table cannot be well described by a straight line, but can be represented by a curve of some kind. In such cases the regression is not linear but curvilinear or in general non-linear. When the regression is non-linear a curve joining the means of successive arrays of the successive columns or rows will fit these mean values more exactly than will a straight line. A non-linear relationship exists between the variables of human age and physical strength. To a certain point, physical strength increases with age, but beyond that point, physical strength begins to decline. A non-linear relationship is also encountered in psychophysics and in experiments dealing with fatigue, practice, learning and forgetting.

The scattergram in Figure 7.16 illustrates as non-linear relationship between the scores on a performance test and chronological age ranging between six and fifteen years Here the score on performance test is time required for completion; thus a high number indicates a poor performance, and the trend is downwards. But the relationship clearly drops most quickly, during the first four years and shows slight variations from year to year during the last four years.

Two regression curves are drawn in the figure to show the trend of relationship, differing in shape as well as in slope. The regression of test score on age is shown by the solid line that is drawn joining the points, which are plotted at the means of the columns. The regression of age upon test score is shown by the dotted line that is drawn joining the points, which are plotted at the means of the rows.

Two correlation ratios are computed for each of the regressions and they will not necessarily be the same in value. The correlation ratio for the regression of X and Y is denoted be η_{XY} (eta-coefficient of X on Y and the.correlation ratio for the regression of Y on X is denoted by η_{XY} (eta-coefficient of Y on X).

The eta coeffiients are not generally used by the beginners in the educational research and for this reasons their calculation is not given here.

USES OF CORRELATION

In addition to the uses described under the different methods of corelation, the coefficient of correlation is used;

(*i*) to evaluate the degree of reliability and validity of psychological tests and inventories;

(*ii*) in the technique of factor analysis; and

(*iii*) to make predictions.

Reliability and Validity of Psychological Test and Inventories

The process of correlation is used in evaluating the different types of reliability. In the test-retest method, the scores on two or more successive administrations of the test are correlated to obtain tests-retest reliability coefficient. The scores on the equivalent (or parallel) forms of the test are correlated in the method of equivalent form reliability. In the split-half method of reliability, the scores on the two halves of the test are correlated, and the Spearman-Brown prophency formula is used to obtain the reliability of the full test.

In the criterion-related validity, tests are usually validated by correlating test scores against some outside criteria. The outside criteria may be scores on tests of accepted validity, ratings by experts, academic achievement, etc. The validity coefficient which is the coefficient of correlation between the test scores and outside criterion, provides a measure of the degree of validity.

Factor Analysis

Factor analysis was originally designed to explore the underlying variables in human abilities which Thurstone has called primary abilities. It has now seen even more use in the search for primary interests, attitudes, and personality traits. Factor analysis attempts to analyse the inter-correlations among the variables so as to reduce them to a smaller number of psychological dimensions which will account for the wide diversity of individual performance. Factor analysis can provide valuable insights into the nature of the tests, which can then be helpful to show as to what extent a particular group of tests are the measures of the same psychological factor and the use of all these tests to measure this factor is unnecessary. This technique, therefore, is helpful in factorial purification of psychological tests and a consequent reduction in the degree of overlapping among them.

Ever since Spearman's discovery in 1904 that coefficients of correlation between tests tend to fall into hierarchical order and his explanation of this on the basis of his famous theory of two factors, one general '*g*' and other specific '*s*', a number of procedures for factor analysis to extract common factors have been proposed. All of them start with a correlation matrix though some of these can also use a matrix of co-variance. The different procedures for factor analysis may be divided into broad groups: bifactor analysis and multifactor analysis.

Holzinger's bifactor method may be looked upon as another natural extension of the simple two-factor plan of analysis. It tries to analyse a battery of tests in such a way that each test yields some relationship to a '*g*' factor and to one, but only one, group factor. The group factors run through matually exclusive groups of tests each, while the '*g*' factor is common to all the tests in the battery. But also gave a method for bifactor analysis in 1940. Commenting on the method of bifactor analysis Guilford (1954, p. 478) says:

> The Burt and Holzinger methods impose some arbitrary restrictions, one of which is requirement of a '*g*' factor. A '*g*' factor can always be found, if one insists upon it. But this theoretical bias may well predispose us to quite misleading results.

Among the chief multi-factor analysis procedures one may include methods of 'principal components' of Hotelling of 'principal axes' of Kelley, of 'maximum likelihood' of Lawley; and the 'Centroid Method' of Thurstone. The first two have much in common. With regard to these Guilford (1954,pp. 477-478) says:

> The methods of Hotelling and Kelley are undoubtedly more rigorous mathematically and can be applied completely objectively. Without modifications, however, they frequently lead to factors by which one can account for obtained scores and intercorrelations but which are difficult to interpret psychologically.

With respect to Hotelling's method of principal component Thomson (1960, p. 119) says:

> The centroids can best be looked at as an approximations to the principal components, more easily calculated.

In connection with maximum likelihood method of Lawley, Thomson (1960, p. 135) says:

> It unfortunately takes, however, much longer to perform than a centroid analysis, particularly when the battery of tests is a large one and when several factors are to be fitted.

In view of the disadvantages of the methods of principal components, principal axes and maximum likelihood with respect either to psychological interpretation or calculations, the researchers invariably use Thurstone's centroid method. In addition to the process of extracting factors, Thurstone, especially has devised rotation methods which he insists are essential for arriving at psychologically meaningful factors. The centroid method has been developed on the basis of matrix algebra which goes beyond the treatment of this book. For the fundamental ideas about the extraction of factors, rotation of reference axes and interpretation of factors the readers are advised to consult Psychometric Methods (Guilford, 1954) and Multiple Factor Analysis (Thurstone, 1947).

Limitations of Factor Analysis

Factor analysis is subject to a number of limitations. They are listed as under:

1. Factor analysis rests on the concept of correlation, with all its inherent inaccuracies and shortcomings. A coefficient of correlation is always relative to the situation under which it is obtained, and its size does not represent a fact in any absolute sense.

2. In research problems it is difficult to determine how many factors are present in a domain in which we are making the analysis. The factors have to be identified on the basis of judgement, and there is no single infallible index or criterion which helps us to determine how many factors should be extracted from a given correlational matrix.

3. Only factors that have been included in the intercorrelation matrix can arise from factor analysis. Any variable that cannot be measured precisely with our present tests and instruments is not likely to emerge as a factor. According to Mouly (1964, p. 308):

> There is no way of determining the factorial structure underlying human behaviour: the only factors that can be discovered are those that have been included in the matrix. Consequently, the factorial pattern which is discovered in a particular study cannot be interpreted as conclusive evidence of the existence or significance of the factors discovered.

4. In order to give psychological meaning and fixity to the factor partern, Thurstone (1947) introduced the idea of rotation to simple structure. However, there is some analysis in which simple structure cannot be achieved or in which more than one solution approaches simple structure.

Prediction

1. Simple prediction. Suppose that in a group of 100 students (Figure 7.10), we wish to estimate a certain boy's height knowing his weight to be 125 pounds. The best possible 'guess' that we can make of this boy's height is the mean height of all of the boys who fall in the 120-129 weight interval. The mean height of the 12 boys in the column of 120-129 weight interval is the most likely height of a boy who weighs 125 pounds. In the same way, the most probable height of a boy who weighs 142 pounds is the mean height of six boys who fall in the weight column 140-149 pounds. And, in general, the most probable height of any boy, in the group is the mean of heights of all boys who weigh the same as he, that is, who fall within the same weight column.

Similarly, we can make best possible guess of the weight of a boy knowing his height. The most probable weight of any boy in the group is the mean weight of all of the boys who are of the same or approximately same height.

It shows that from the scattergram alone it is possible to 'predict' one variable from another—with the help of column means. But the prediction is rough, and is obviously subject to a large error of estimate. Moreover, in such a prediction we have made no use of our knowledge concerning the overall correlation between the two variables of height and weight. The angle or the degree of separation of the regression lines in Figure 7.11 indicates the size of the correlation coefficient and therefore, describes the relationship between the variables of height and, weight over the whole range. The knowledge of the equations of the regression lines is necessary if we are to make a prediction based upon all of our data.

The equations of the two regression lines in a correlation table represent the straight lines which best fit the means of the successive columns and rows in the scattergram. More precisely, the line of regression is the straight line which gives the 'best fit' in the least square sense, to the given frequency distribution. If the straight line is so chosen that the sum of squares of deviations parallel to the axis of Y is minimized, it is called the line of regression of Y on X and it gives the best estimate of Y for any given value of X.

Let $y=a+bx$ be the equation of the line of regression of Y on X, so that corresponding to $x=x_i$, the expected value of y is Y_i $(=a+bx_i)$ and S, the sum of squares of deviations of observed

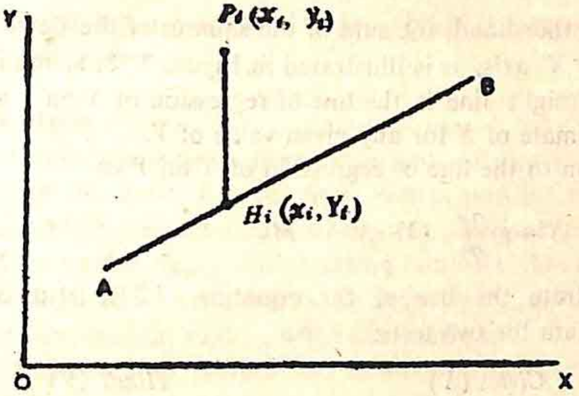


FIG. 7.17. Regression Line of *Y* on *X*.

values y_i from the expected value Y_i, is given by

$$S=\sum_{i=1}^{n} f_i\,(y_i-Y_i)^2$$

$$=\sum_{i=1}^{n} f_i\,[y_i-(a+bx_i)]^2$$

According to the principle of least squares we have to choose a and b so that S is minimum.

By a mathematical procedure (which is beyond the scope of this book), we find the values of a and b so that the equation to the line of regression of Y on X is:

$$Y'=r\frac{\sigma_y}{\sigma_x}(X-M_x)+M_y \qquad (7.27)$$

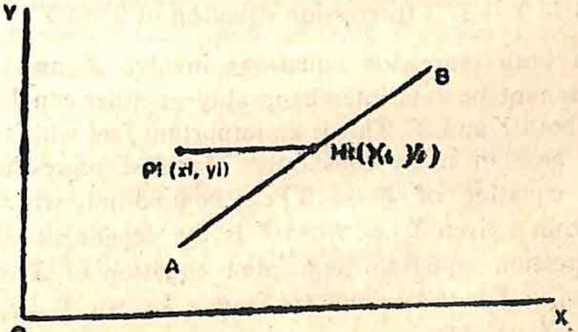


FIG. 7.18. Regression Line of *X* on *Y*.

If, on the other hand, the sum of the squares of the deviations parallel to the X axis, as is illustrated in Figure 7.18, is minimised, the resulting straight line is the line of regression of X on Y and it gives the best estimate of X for any given value of Y.

The equation to the line of regression of X on Y is:

$$X' = r\frac{\sigma_x}{\sigma_y}(Y - M_y) + M_x \tag{7.28}$$

To illustrate the use of the equation (7.27), let us consider the following data for two tests:

Civics (X)	*Hindi* (Y)
$M_x = 75.00$	$M_y = 70.00$
$\sigma_x = 6.00$	$\sigma_y = 8.00$

$r_{xy} = 0.72$

Suppose a student scores 65 marks in the test of civics and we wish to estimate his most probable score in the test of Hindi. Substituting the given values from the data in the equation (7.27), $Y' = 60$; and accordingly, the most probable score in Hindi of a student with score of 65 in Civics is 60.

If the problem is to predict marks in civics instead of Hindi, we must use the second regression equation (7.28). Now, if a student scores 50 marks in Hindi, we find, on replacing Y by 50 in the equation (7.28), that $X' = 64$. Hence, the most probable score in civics of a student who has scored 50 in Hindi is 64.

The two regression equations worked out from the above data are:

$Y' = 0.96\,X - 2$ (regression equation of Y on X).

and $X' = 0.54\,Y + 37.2$ (regression equation of X on Y).

Although both regression equations involve X and Y, the two equations cannot be used interchangeably—neither can be employed to predict both Y and X. This is an important fact which the researcher must bear in mind constantly. The first regression equation (regression equation of Y on X) can be used only when Y is to be predicted from a given X i.e., when Y is the dependent variable. The second regression equation (regression equation of X on Y) can be used only when X is to be predicted from a known Y i.e. when X is the dependent variable.

In summary, there are two regression lines represented by two equations in a scattergram, the one through the means of the columns and the other through the means of rows. This is always true unless the correlation is perfect.

Errors in Prediction

When the correlation between the two variables X and Y is perfect, there would be no error of prediction. Perfect positive correlation indicates that for every increase in X, there was a proportional increase in Y, and the perfect negative correlation indicates that for every increase in X, there was a proportional decrease in Y. But when the correlation is not perfect, that is, when the coefficient of correlation is less than $+1.00$ or -1.00, there would be error of prediction.

A measure for estimating this error of prediction is known as the standard error of estimate and is denoted by $\sigma_{e}st$.

The formula for the standard error of estimate when Y measures are predicted from X measures is:

$$\sigma_{e}sty=\sigma_{y}\sqrt{1-r^{2}} \qquad (7.29)$$

in which

σ_{y}=the *SD* of the Y distribution

r=the coefficient of correlation between X and Y

From the regression equation of Y on X, we computed the most probable score of a student in Hindi to be 60 when his score in Civics was 6.5. In order to determine the accuracy of this prediction, we make use of the formula (7.29) to get:

$$\sigma_{e}sty=8\sqrt{1-(.72)^{2}}$$
$$=5.5$$

We can now say that the most probable score of a student in Hindi scoring 65 in Civics is 60 with a standard error of estimate of 5.5.

When the X measures are predicted from Y measures, the standard error of estimate is:

$$\sigma_{e}stx=\sigma x\sqrt{1-x^{2}} \qquad (7.30)$$

In the previous example, we estimated from the regression equation of X on Y that the most probable score of a student in civics scoring 50 marks in Hindi was 64.

Using formula (7.30), the $\sigma_{e}stx$ of this prediction is:

$$\sigma_{e}stx=6\sqrt{1-(.72)^{2}}$$
$$=4.16$$

Hence, our best estimate of the score in Civics of any student in the group of students has a standard error of 4.16.

2. *Multiple prediction.* In multiple prediction, the estimate of the dependent variable is based on the linear combination of several independent variables. It makes use of multiple regression equation. For the three-variable problem, the general form of the regression equation is

$$X'_1=a+b_{12\cdot3}\,X_2+b_{13\cdot2}\,X_3 \tag{7.31}$$

in which the dependent variable X_1 is predicted from the independent variables X_2 and X_3. The $b_{12\cdot3}$ and $b_{13\cdot2}$ are known as partial regression coefficients and are given by the formula:

$$\begin{aligned} b_{12\cdot3}&=\left(\frac{\sigma_1}{\sigma_2}\right)\beta_{12\cdot3}\\ b_{13\cdot2}&=\left(\frac{\sigma_1}{\sigma_2}\right)\beta_{13\cdot2} \end{aligned} \tag{7.32}$$

The betas are found by the formulae:

$$\begin{aligned} \beta_{12\cdot3}&=\frac{r_{12}-r_{13}\,r_{23}}{1-r^2_{23}}\\ \beta_{13\cdot2}&=\frac{r_{13}-r_{12}\,r_{23}}{1-r^2_{23}} \end{aligned} \tag{7.33}$$

Making use of data in Table 7.20 we compute

$$\beta_{12\cdot3}=\frac{.54-.37\times .26}{1-(.26)^2}=0.48$$

$$\beta_{13\cdot2}=\frac{.37-.54\times .26}{1-(.26)^2}=0.25$$

$$b_{12\cdot3}=\frac{19.64}{2.89}\times 0.48=3.26$$

$$b_{13\cdot2}=\frac{19.64}{7.05}\times 0.25=0.70$$

$$\begin{aligned} a&=119.73-(3.26\times 11.26)-(0.70\times 7.05)\\ &=78\text{-}09 \end{aligned}$$

Thus the regression equation comes out to be

$$X'_1=78.09+3.26X_2+0.70X_3$$

With more than three variables, we make use of method of Pivotal Condensation, originated by Aitken (1937).

To illustrate application of this method we take up a problem with five variables, one criterion or dependent variable (X_1) and four predictors or independent variables (X_2, X_3, X_4 and X_5). The criterion variable is the academic achievement and the predictors are psychological variables of intelligence, creative thinking, achievement-motivation and socio-economic status. The intercorrelations among the five variables alongwith means and standard deviations are shown in Table 7.21.

TABLE 7.21 : Correlation Coefficients between Criterion and four Predictors Alongwith Means and Standard Deviations

	X_1	X_2	X_3	X_4	X_5
X_1	1.00	.548	.304	.199	.252
X_2	.548	1.00	.319	.277	.355
X_3	.304	.319	1.00	.394	.377
X_4	.199	.277	.394	1.00	.451
X_5	.252	.355	.377	.451	1.00
M	433.94	43.28	100.54	13.58	34.89
SD	103.13	14.07	18.31	6.92	12.59

The procedure, as explained by Aitken, requires the successive calculation of differences between cross products. If the four cell values are:

$$\begin{matrix} a & b \\ c & d \end{matrix}$$

the difference between cross products is ad-bc. In this case the cell value '*a*' is the pivotal element.

The steps followed in the calculation are listed as under alongwith computations which are summed up in Table 7.22.

1. Write down the matrix of correlations among the predictors (X_2, X_3, X_4 and X_5) and insert. Is along the diagonal. Write a row containing the correlations of the predictors with criterion below this matrix. The resulting matrix is shown to the right of slab *A* in Table 7.22. To the right of this matrix record another matrix with four entries, that is,—Is along the diagonals and all other elements including those in the bottom row to be *O*.

2. Sum the rows to obtain the values in the check column.

3. Calculate the differences between cross products for the first two rows of slab *A*, using the (1) in the top left cell as the pivotal element. The following product differences are formed:

$$(1)\,(1)-(.319)\,(.319)\;=.898$$
$$(1)\,(.394)-(.277)\,(.319)=.306$$
$$(1)\,(.377)-(.355)\,(.319)=.264$$
$$(1)\,(0)-(-1)\,(.319)\;=.319$$
$$(1)\,(-1)-(0)\,(.319)\;=-1$$

$(1)(0)-(0)(.319) = 0$
$(1)(0)-(0)(.319) = 0$

These values are entered in the first row of slab B. The check value is obtained by forming the product difference: $(1)(1.090)-(.951)(.319)=.787$; which should be equal to the sum of elements in the first row of slab B.

4. The second row of slab B is obtained by dividing each element by the top left element, *i.e.*, .898. The result is a row with unity as the pivot. This helps subsequent computations. This part of the procedure may be accomplished by multiplying the element in the row by the reciprocal of .898, *i.e.*, by 1.114. The remaining elements in slab B are obtained by forming the product differences using first row of slab A with third, fourth and fifth rows of slab A, successively, always using the (1) in the top left cell as the pivotal element. We can have:

for 3rd row:

$(1)(.394)-(.319)(.277)=.306$
$(1)(1)-(.277)(.277)=.923$
....
....
....
$(1)(0)-(0)(.277) = 0$

for 4th row:

$(1)(.377)-(.319)(.355)=.264$
$(1)(.456)-(.277)(.355)=.353$
....
....
$(1)(-1)-(0)(.355) = -1$

and for 5th row

$(1)(.304)-(.319)(.548)=.129$
$(1)(.199)-(.277)(.548)=.047$
....
....
$(1)(0)-(0)(.548) = 0$

Each row is summed to provide a check on the calculation.

5. The procedure is now repeated to obtain slabs C, D and E. At each stage, with the exception of the last, the top row in each slab is divided by the left hand cell value, or multiplied by the reciprocal of that value, to obtain a second version of the top row. The appropriate reciprocal for row C is 1.221 and 1.405 for row D.

TABLE 7.22: Aitken's Method for Computing Regression Coefficients

									Check
A	(1)	(.319)	.277	.355	—1	0	0	0	0.951
	.319	(1)	.394	.377	0	—1	0	0	1.090
	.277	.394	(1)	.451	0	0	—1	0	1.122
	.355	.377	.451	(1)	0	0	0	—1	1.183
	.548	.304	.199	.252	0	0	0	0	1.303
	(1.114)	(.898)	(.306)	.264	.319	—1	0	0	.787
B	1.00	.341	.294	.355	—1	.114	0	0	.876
	.306	.923	.353	.277	0		—1	0	.859
	.264	.353	.874	.355	0		0	—1	.846
	.129	.047	.057	.548	0		0	0	.781
	(1.221)	.819	.263	.168	.341		—1	0	.591
C		1.00	.321	.205	.416	—1	.221	0	.721
		.263	.796	.261	.294	0		—1	.614
		.003	.019	.502	.144	0		0	.668
	(1.405)		.712	.207	.185		.321	—1	.425
D			1.00	.291	.260		.451	—1.405	.597
			.018	.501	.143		.004	.000	.666
E	Regression Coefficients :			0.496	0.138		—0.004	0.025	0.655

The four cell values in slab E are the condensation of the original matrix. These are the multiple regression coefficients for predicting the criterion from the four predictors.

By multiplying the correlations of predictors with the criterion by the corresponding regression coefficients, summing the resulting products and taking the square root, we obtain tbe multiple correlation coefficient (R) as follows:

$$R = \sqrt{.548 \times (.496) + .304 \times (.138) + .304(-.004) + .252 \times (.025)}$$
$$= 0.565$$

The significance of the multiple correlation coefficient is tested by computing F value which is given by the following formula

$$F = \frac{R^2/g}{(1\text{-}R^2)\ (N\text{-}g\text{-}1)} \qquad (7.34)$$

where

R=multiple correlation coefficient.
g=number of independent variables
N=number of observations

The Table D is entered with $df_1=g$ and $df_2=N\text{-}g\text{-}1$,

The regression equation for predicting the criterion variable from four predictors is:

$$X'_1=a+\beta_2\frac{\sigma 1}{\sigma 2}X_2+\beta_2\ \frac{\sigma 1}{\sigma 3}X_3+\beta_4\ \frac{\sigma 1}{\sigma_4}X_4+\beta_5\ \frac{\sigma 1}{\sigma_5}X_5$$

$$a=\overline{X}_1-\beta_2\ \frac{\sigma 1}{2}\overline{X}_2-\beta_3\frac{\sigma 1}{3}\ \overline{X}_3-\beta_4\ \frac{\sigma 1}{4}\overline{X}_4+\beta_5\ \frac{\sigma 1}{5}\ \overline{X}_5 \quad (7.35)$$

in which

β_2, β_3, β_4 and β_5 are the regression coefficients of the predictor variables and σ_1, σ_2, σ_3, σ_4, and σ_5 are the standard deviations of the respective variables, under study.

Thus the multiple regression equation may be put as under:

$$X_1=a+(.496)\frac{103.13}{14.07}X_2+(.138)\ \frac{103.13}{18.31}X_3$$
$$+(-.004)\ \frac{103.13}{6.92}X_4+(.025)\frac{103.13}{12.59}X_5$$

and

$$a=433.94-(.496)\frac{103.13}{14.07}(43.28)-(.138)\frac{103.13}{18.31}(100.54)$$
$$-(-.004)\frac{103.13}{6.92}(13.58)-(.025)\frac{103.13}{12.59}(34.89)$$

while solving the equation we get

$X'_1=192.12+3.636X_2+.777X_3-.060X_4+.205X_5$.

The calculation of multiple regression weights is clearly a laborious procedure. With any increase in number of variables, the use of Wherry-Doolittle method, which is beyond the scope of this book, for selecting significant contributors from amongst the given set of predictors is suggested.

LIMITATIONS OF STATISTICAL METHODS

Though statistics is an important tool of a researcher, there are a number of limitations that should be recognized in using statistical methods, and in drawing conclusions from statistical treatment of the data. Best (1977, p. 261) has listed them as under:

1. Statistical process is the servant of logic and only has value if it verifies, clarifies, and measures relationships that have been established by clear, logical analysis. Statistics is a mean, never an end, of research.
2. A statistical process should not be employed in the analysis of data unless the basic assumptions and limitations underlying its uses are clearly understood.
3. The conclusions derived from statistical analysis will be no more accurate or valid than the original data. To use an analogy, no matter how elaborate the mixer, a cake made of poor ingredients will be a poor cake. All the refinement of elaborate statistical manipulation will not yield significant truths if the data results from crude or inexact measurement.
4. All treatment of data must be checked and double checked frequently to minimize the likelihood of errors in measurement, recording, tabulation and analysis.
5. There is a constant margin of error wherever measurement by human beings is involved. This error is increased when qualities or characteristics of human personality are subjected to measurement, or when inferences about the population are made from measurements derived from statistical samples.

   When comparisons or contrasts are made, a mere number difference is in itself, not a valid basis for any conclusion. A test of statistical significance should be employed to weigh the possibility that chance in sample selection could have yielded the apparent difference. To apply these measures of statistical significance is to remove some of the doubt from the conclusions.
6. Statisticians and liars are often equated in humorous equips. There is little doubt that statistical processes can be used to prove nearly anything that one sets out to prove. Starting with false assumptions, using inappropriate procedures, or omitting relevant data, the biased investigator can arrive at false conclusions. These conclusions are often particularly dangerous because of the authenticity that the statistical treatment seems to confer.

Statistical techniques, if rightly employed, enable the researcher to analyse the data meaningfully and exactly; to summarize the results in meaningful and convenient form; and to draw general conclusions. No statistical method should be applied unless it adds clarity or meaning to the analysis and treatment of data.

POST TEST

1. Define the term measurement.
2. List and describe some important properties of numbers.
3. List and describe the four levels of measurement scales.
4. Define parametric and non-parametric data.
5. Tabulate the following data into two frequency distribution, using (*a*) an interval of three and (*b*) an interval of five units. Let the first interval begin with score 50.

62	65	67	57	62
71	68	55	76	63
57	72	66	66	60
73	61	53	62	62
51	57	74	59	54

6. Compute the cumulative frequencies and cumulative per cents for the distributions obtained under the problem 5.
7. List the four important methods of representing a data graphically.
8. Construct the histogram or column diagram, frequency polygon, cumulative frequency curve and ogive for the following distribution:

Scores	*f*
90-99	5
80-89	9
70-79	14
60-69	19
50-59	21
40-49	30
30-39	20
20-29	15
10-19	8
0-9	6

9. List the four descriptive statistical measures.
10. List the three measures of central tendency.
11. Define the mean, the median and the mode.
12. Calculate the mean, the median and the mode for the following data:
78. 76, 75, 70, 70, 67, 65, 61, 55, 50, and 48.
13. Calculate the mean, the median and the mode of the distribution given under the problem 8.
14. Describe the uses of the mean, the median and the mode.
15. List the three important measures of spread, dispersion or variability.
16. Define the range, the variance and the standard deviation.
17. Calculate the range, the variance and the standard deviation of the following data:
37, 37, 35, 34, 30, 28, 26, 26, 24, 22, 20, 17.
18. Calculate the range, the variance and the standard deviation of the distribution given under the problem 8.
19. Describe the uses of the range and the standard deviation.
20. Describe the nature and characteristics of the normal probability curve.
21. In a sample of 500 cases the mean of a certain test is 15.20 and standard deviation is 2.80. Assuming normality of distribution.
    (*a*) How many individuals score between 11 and 15?
    (*b*) How many score above 20? Below 10?

(c) What are the chances that any individuals selected at random will score above 16?

22. Describe the applications of the normal curve in the field of educational research.
23. List the four measures of comparing individuals on the basis of different types of scores.
24. Define the sigma scores (σ scores) standard scores (Z or T scores), percentiles and percentile ranks.
25. Find the sigma scores corresponding to the raw scores of 56 and 67 in a distribution with a mean of 60 and standard deviation of 5.
26. Given a distribution with mean=80 and $\sigma=12$. Shyam's score is 85 and Ram's, 76. Express these raw scores as standard scores in a distribution with a mean of 500 and standard deviation of 100.
27. Given a distribution with mean=90 and $\sigma=15$, Prakash's score is 75 and Gurmeet's 86. Express these raw scores as the T scores.
28. Define a percentile.
29. Calculate p_{99}, p_{90}, p_{70}, p_{20}, and p_{10} for the following distribution:

Scores	f
152-155	8
148-151	5
144-147	12
140-143	58
136-139	40
132-135	22
128-131	12
124-127	15
120-123	10
116-119	3

30. Define the percentile ranks.
31. Kailash ranks 8th in a class of 40 in Hindi and 8th in a class of 30 in Mathematics. Compare his percentile ranks in the two subjects.
32. The scores given below were achieved upon a group intelligence test (X) and algebra test (Y) by 50 students of seventh grade in a certain school. Construct a scattergram for these data:

(X)	(Y)	(X)	(Y)	(X)	(Y)
48	173	47	153	53	143
38	134	39	145	46	173
26	179	32	134	39	168
37	159	37	184	26	164
34	167	26	154	33	127
51	136	40	90	44	144
46	152	31	96		

35	160	50	159	35	160
49	106	29	175	48	149
40	95	41	164	40	149
57	146	32	111	43	143
23	175	50	164	38	159
51	126	29	119	37	157
35	120	41	160	40	120
41	154	32	149	36	140
28	146	49	149	43	141
32	154	58	143	48	143

33. Plot the regression lines with the help of the scattergram obtained under the problem 32.
34. List the various measures of relationship.
35. Calculate product moment correlation between the following two sets of scores:

X	*Y*
25	22
28	36
32	24
27	31
29	18
30	20
26	16
31	15
25	23
38	29

36. Calculate the product moment correlation in the scattergram obtained under the problem 32.
37. Calculate the rank order coefficient of correlation between the two sets of scores given below:

Subjects	*X*	*Y*
A	10	8
B	13	15
C	17	11
D	12	10
E	14	12
F	15	17
G	11	10
H	16	8
I	11	14
J	17	16

38. Calculate the biserial correlation and the point serial correlation for the following distribution. Which is more appropriate for these data?

Scores on music appreciation test	*Groups with training in music* *f*	*Groups without training in music* *f*
80-84	4	12
75-79	2	16
70-74	5	20
65-69	5	25
60-64	1	19
55-59	2	10
50-54	0	8
45-49	1	2

39. Two items X and Y are part of a test of 100 items. Item X is passed by 90 and failed by 110 students in a group of 200. Item Y is passed by 130 and failed by 70 in the same sample. Find the tetrachoric correlation and the phi coefficient correlation between X and Y. Which is the more appropriate for these data?

		Item X		
		Failed	*Passed*	
Item Y	Passed	70 B	60 A	130 A+B
	Failed	40 D	30 C	70 C+D
		110	90	

40. Calculate the coefficient of contingency for the following table given below:

		Father's Eye Colour				
		Blue	*Grey*	*Hazel*	*Brown*	*Total*
Son's Eye Colour	Blue	184	80	46	40	350
	Grey	73	134	46	46	299
	Hazel	15	44	60	33	152
	Brown	46	46	48	119	259
	Total	318	304	200	238	1060

41. The correlation between a general intelligence test and achievement in mathematics in a group of children from 10 to 15 years old is 0.85. The

correlation between the general intelligence test and age is 0-65; and the correlation between achievement in mathematics and age is 0.55. What is the correlation betweenachievement in mathemetics and general intelligence in children of the same age?

42. Let X_1 be a criterion and X_2 and X_3 be two other tests. the correlations, means and σ's are as tollows:

$r_{12}=.55$ $\sigma_1=5.00$ $M_1=70.00$
$r_{13}=.45$ $\sigma_2=.800$ $M_2=80.00$
$r_{23}=.20$ $\sigma_3=7.00$ $M_3=30.00$

Calculate R_1 (23).

43. Define correlation ratio.
44. Describe the uses of various methods of correlation.
45. Describe the various of correlation.
46. Describe the nature and uses of factor analysis.
47. List certain important methods of factor analysis.
48. Describe the various limitations of factor analysis.
49. Describe the nature of simple aud multiple prediction.
50. Write down the regression equations ot (*a*) X on Y and (*b*) Y on X.
51. Define the standard error of estimate.
52. Write down the formulae for the standard error of estimate when (*a*) Y measures are predicted from X measures and (*b*) X measures are predicted from Y measures.
53. Given the following data for two tests:

Algebra (X)	*Physics (Y)*
$M_x=70.00$	$My=65.00$
$x=5.00$	$\sigma y=7.00$

$r_{xy}=0.80$

(*a*) Work out the regression equations.
(*b*) Predict the probable grade in Physics of a student whose Algebra mark is 60. Find σ_{est} of this prediction.
(*c*) Predict the probable grade in Algebra of a student whose Physics mark is 55. Find σ_{est} of this prediction.

54. Write down the multiple regression equation involving three variables.
55. Write down the formulae for calculating partial regression coefficients and betas.
56. List and describe briefly the limitations of statistical methods.

8. Inferential Statistics

This chapter has two major sections. The first section deals with the statistical inference based on parametric tests. It includes the assumptions on which the use of parametric tests are based; inference regarding means of large and small samples; significance of the difference between the means in large and small independent samples; significance of the difference between the means in large and small dependent samples; analysis of variance; analysis of co-variance; significance of Pearson's coefficient of correlation; and significance of the difference between Pearson's coefficient of correlations. The second section deals with the statistical inference based on non-parametric tests. It includes the assumptions on which the use of non-parametric tests are based; the use of chi-square test, the median test, the Mann-Whitney U test, the Kolmogorov-Smirnov test, the Wilcoxon matched-pairs signed-ranks test, the McNemar test for significance of changes; the significance of Spearman's coefficient of co-relation; significance of Phi-coefficient of correlation and significance of contingency coefficient of correlation.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. List and describe the assumptions on which the use of parameteric tests are based.
2. Define the sampling distribution of means.
3. State the characteristics of Central Limit Theorem.
4. Define the standard error of mean.
5. Define the confidence intervals and levels of confidence.
6. Compute the .95 and .99 confidence intervals for the true mean from the large sample mean.
7. Define and illustrate the concept of degrees of freedom.
8. Compute the .95 and .99 confidence intervals for the true mean from the small sample mean.
9. Test the significance of the difference between the means of two independent samples involving: (*a*) one tailed test and (*b*) two tailed test.
10. Test the significance of the difference between the means of two independent small samples involving: (*a*) one tailed test and (*b*) two tailed test.
11. Test the significance of the difference between the means of two dependent large samples.
12. Test the significance of the difference between means of two dependent small samples.
13. Describe the uses of analysis of variance.

14. List and state the basic assumptions underlying the technique of analysis of variance.
15. Test the basic assumptions underlying the technique of analysis of variance.
16. Apply analysis of variance to test the significance of the difference between the means involving a randomized group design.
17. Apply analysis of variance to test the significance of the difference between the means involving a factorial design.
18. Describe the uses of analysis of co-variance.
19. List and state the basic assumptions underlying the technique of analysis of co-variance.
20. Test the basic assumptions underlying the technique of analysis of co-variance.
21. Apply analysis of co-variance to test the significance of the difference between the means.
22. Compute the .95 and .99 confidence intervals for the true coefficient of correlation from the sample Pearson's coefficient of correlation.
23. Test the significance of the difference between Pearson's coefficient of correlations.
24. List and describe the assumptions on which the use of non-parametric tests are based.
25. Describe the use of chi square test.
26. Illustrate the use of chi square test.
27. Describe the use of median test.
28. Illustrate the use of median test.
29. Describe the use of Mann-Whitney U test.
30. Illustrate the use of Mann-Whitney U test.
31. Describe the use of Kolmogorov-Smirnov two sample test.
32. Illustrate the use of Kolmogorov-Smirnov test when applied to (*a*) small samples and (*b*) large samples.
33. Describe the use of sign test.
34. Illustrate the use of sign test when applied to (*a*) small samples and (*b*) large samples.
35. Describe the use of Wilcoxon matched-pairs signed-ranks test.
36. Illustrate the use of Wilcoxon matched pairs signed-ranks test when applied to: (*a*) small samples and (*b*) large samples.
37. Describe the use of McNemar test for significance of changes.
38. Illustrate the use of McNemar test.
39. Test the significance of Spearman's coefficient of correlation when (*a*) the size of the sample is small and (*b*) when the size of the sample is large
40. Test the significance of phi-coefficient of correlation.
41. Test the significance of contingency coefficient of correlation.

So far we have been concerned primarily with descriptive statistics. The computed values were used to describe the properties of particular samples. Now we shall turn from descriptive statistics to inferential or sampling statistics, which are useful because they enable the researcher to make generalizations or inferences about populations from his observations of the characteristics of samples. Research is generally conducted by means of a sample on the basis of which generalizations concerning the population from which the sample was drawn are reached. Several sampling methods were described earlier in Chapter Five, each one particularly appropriate in a given set of conditions. The researcher computes certain statistics (sample values) as the basis for inferring what the corresponding parameters (population values) might be as it is rarely, if ever, possible to measure all of the members of a given population. The value of the parameters for a given population are unknown.

Ordinarily, the researcher draws a single sample from a given population and his problem becomes one of determining how well he can infer or estimate the parameter from the one sample statistics. But he can, under specified conditions, forecast the parameters from the sample statistics with a known degree of accuracy. The degree to which a sample statistics represents its parameter is an index of the significance of the computed sample statistics. While it is possible to make inferences about various parameters, in this chapter we shall limit our discussion to mean and correlation coefficient. For making the inferences about various parameters, the researcher makes use of parametric and non-parametric tests.

STATISTICAL INFERENCE BASED ON PARAMETRIC TESTS

Parametric tests are useful and are the most powerful tests for testing the significance or trustworthiness of the computed sample statistics. Their use is based upon certain assumptions. These tests should be used:

1. When the variables described are expressed in interval or ratio scales and not in nominal or ordinal scales of measurement.
2. When the population values are normally distributed.
3. When the samples have equal or nearly equal variances. This condition is known as equality or homogeneity of variances and is particularly important to determine when the samples are small.
4. When the selection of one case in the sample is not dependent upon the selection of any other.

Inferences Regarding Means of Large Samples

Suppose that we wish to measure the general intelligence of the tribal high school adolescents of Himachal Pradesh by means of a non-verbal test of intelligence developed for tribal adolescents of this state. It would be a difficult if not an impossible task to measure the general intelligence of all of the tribal high school adolescents of the state and hence we must usually be satisfied with a sample drawn from this population. This sample should be as large and as randomly drawn as possible to represent adequately all the tribal high school adolescents of the state. If we select a large number of random samples of 100 tribal adolescents each from the population of all tribal high school adolescents in Himachal Pradesh, the mean values of general intelligence for all samples would not be identical. A few would be relatively high, a few relatively low, but most of them would tend to cluster around the population mean. This variation of sample means is due to what is known as sampling error. The sample mean will not only vary from sample to sample but will also usually deviate from the population mean. Each of these sample means can be treated as a single observation and we can set up their frequency distribution. This distribution is known as sampling distribution of the means.

According to the Central Limit Theorem, if large-sized samples, greater than 30 in size, are selected at random from an infinite population:

1. The distribution of the sample means is normal and it possesses all the characteristics of a normal distribution.
2. The average value of the sample means will be the same as the mean of the population.
3. The distribution of the sample means around the population mean has its own standard deviation. This standard deviation is known as the standard error of the mean and is denoted as SE_M or σ_M.

The standard error of the mean gives us a clue as to how far such sample means may be expected to deviate from the population mean. If the mean of a particular sample may be taken as an estimate of the population mean, any deviation of such a sample mean from the population mean may be regarded as an error of estimation. The standard error of a mean tells us how large the errors of estimation are in any particular sampling situation.

The formula for the standard error of the mean in a large sample is:

$$SE_M \text{ or } \sigma_M = \frac{\sigma}{\sqrt{N}} \tag{8.1}$$

in which

σ=the standard deviation of the population
and N=the number of cases in the sample.

This formula requires the knowledge of the parameter (i.e. standard deviation of the population) in order to compute the standard error of the mean. Since the value of $\bar{\sigma}$ is not known, we have to make the estimate of this standard error.

When we have computed mean (M) and standard deviation (σ) of a large sample, the estimation of $\bar{\sigma}_M$ can be made by the formula

$$\bar{\sigma}_M = \frac{\sigma}{\sqrt{N}} \qquad (8.2)$$

in which

σ=standard deviation of sample

N=the number of cases in the sample

To illustrate the use of the formula (8.2), let us assume that the mean of the intelligence test scores of a sample of 100 tribal high school adolescents is 30 and the standard deviation 5. The standard error of the mean is:

$$SE_M = \bar{\sigma}_M = \frac{5}{\sqrt{100}} = 0.50$$

This SE_M can be thought of as the standard deviation of a distribution of sample means around the fixed population mean of all tribal high school adolescents. In case of large samples, the sampling distribution of sample means is assumed to be normal.

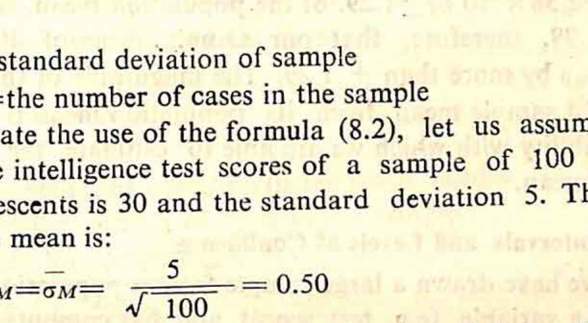


FIG. 8.1. Sampling Distribution of Means showing Variability of obtained Means around Population Mean in terms of σ_M.

The normal curve in Figure 8.1 illustrates that this sampling distribution is centred at the unknown population mean with its standard deviation 0.50. The sample means fall equally often on the positive and negative sides of the population mean. About 2/3 of our sample means (exactly 68.26 per cent) will lie within $\pm 1.00\ \sigma_M$ of the population mean, i.e., within a range of $\pm 1 \times 0.50$ or $\pm$ 0.50. Furthermore, 95 of our 100 sample means will lie within $\pm 2.00\ \sigma_M$ (more exactly $\pm$ 1.96 σ_M) of the population mean i.e. 95 of 100 sample means will lie within $\pm 1.96 \times 0.50$ or ± 0.98 of the population mean. The probability (P) is .95, therefore, that our sample mean of 30 does not miss the population mean (M_{pop}) by more than ± 0.98; and the probability is .05 that the sample means 30 does miss the M_{pop} by more than ± 0.98 Also ,99 of our sample means will be within $\pm$ 3.00 σ_M (more exactly $\pm$ 2.58 σ_M) of the population mean i.e., 99 of our sample means will lie within $\pm 2.58 \times .50$ or ± 1.29. of the population mean. The probability (P) is .99, therefore, that our sample mean of 30 does not miss the M_{pop} by more than $\pm$ 1.29. The magnitude of the probable deviation of a sample mean from its population mean is a measure of the probability with which we are able to estimate the M_{pop} from the sample mean.

Confidence Intervals and Levels of Confidence

Suppose we have drawn a large sample from a population to obtain measures of a variable (e.g. test score) and has computed the mean M for the sample. Using the central limit theorem and the normal-curve concept, we can say that the sample M has a 95 per cent chance of being within 1.96 standard error units M_{pop}. In other words, a mean for a random sample has a 95 per cent chance of being within 1.96 σ_M units from M_{pop}. It may also be said that there is a 99 per cent chance that the sample mean lies within 2.58 σ_M units of M_{pop}. More specifically, we can state that there is a 95 per cent probability that the limits $M \pm 1.96$ σ_M spans the population mean M_{pop}. By the same reasoning we can say that there is a 99 per cent probability that the limits $M \pm 2.58$ σ_M encloses M_{pop}. Such limits are called as confidence intervals. These limits are also called as fiduciary limits and the confidence placed in the interval defined by the limits as containing the M_{pop} is called the fiduciary probability (Fisher, 1935). These limits help us to adopt particular two levels of confidence. One is known as 5 per cent level, or .05 level, and the other as the 1 per cent or.01 level. The .05 level of confidence indicates that the probability is .95 that M_{pop} lies within the interval $M \pm 1.96\sigma_M$, and .05 that it falls out-side of these limits. By the same reasoning, .01 level of confidence indicates that the probability is .99 that M_{pop} lies within the interval $M \pm 2.58$ σ_M, and .01 that it falls outside of these limits.

By way of illustration, let us apply the concept to the previous problem. Taking as our limits $M \pm 1.96$ σ_M, we have $30 \pm 1.96 \times 0.50$ or a confidence interval marked off by the limits 29.02 and 30.98. Our confidence that this interval contains M_{pop} is expressed by a probability of .95. If we want a higher degree of confidence, we can take the

.99 level of confidence, for which the limits are $M \pm 2.58\ \sigma_M$ or a confidence interval given by the limits 28.71 and 31.29. We may be quite confident that M_{pop} is not lower than 28.71 nor higher than 31.29. i.e., the chances are 99 in 100 that the M_{pop} lies between 28.71 and 31.29.

Inferences Regarding Means of Small Samples

When the number of cases in the sample is less than 30, we may estimate the value of $\overline{\sigma_M}$ by the formula

$$SE_M = \frac{S}{\sqrt{N}} \tag{8.3}$$

in which

S=standard deviation of the small sample

N=the number of cases in the sample.

The formula for computing S is

$$S = \sqrt{\frac{\Sigma_x^2}{N-1}} \tag{8.4}$$

in which

Σ_x^2=sum of the squares of deviations in the sample

N=number of cases in sample.

For small samples, we should not assume that the sampling distribution of means is distributed normally. It was in about 1815 when William Seely Gosset, a consulting statistician for Guinness Breweries of Dublin, Ireland developed the concept of small sample size. He found that the distribution curves of small sample means were somewhat different from the normal curve. Such a distribution was named as t-distribution. When size of the sample is small, the t-distribution

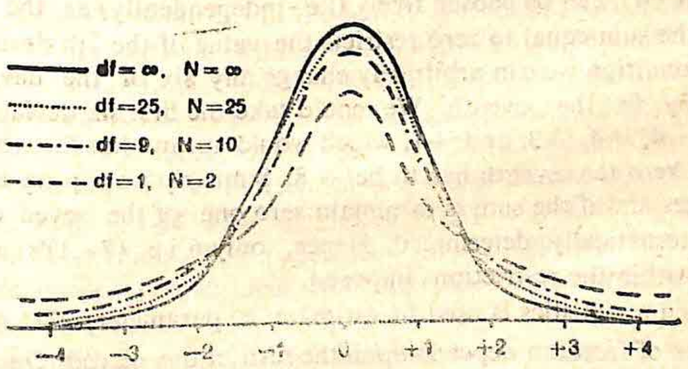


FIG. 8.2. Distribution of t for Degree of Freedom from 1 to ∞ (**When df is very large, the distribution of t approaches normal).**

lies under the normal curve, but the tails or ends of the curve are higher than the corresponding parts of the normal curve. Figure 8.2 shows that the *t*-distribution does not differ greatly from the normal unless the sample size is quite small; and that as sample increases in size the *t*-distribution approaches more and more closely to the normal curve.

For small samples it is necessary to make use of selected points reproduced in table of *t*, Table *C* in the Appendix. As the sample size increases, the *t* values approach the *Z* values of the normal probability table. In the use of *t* values we make use of important concept of *degree of freedom* and it is necessary to introduce the reader to this concept.

Degrees of Freedom

For finding the standard deviation of small samples we make use of $N—1$ in the denominator in the formula (8.4) in place of *N* used in the basic formula (7.5a) for standard deviation. The difference between the two formulae may seem very slight, especially whan *N* is reasonably large, but there is a very important difference in the meaning. In the formula (8.4), $N—1$ is known as the number of degrees of freedom, and is denoted by the symbol *df*. By freedom we mean freedom to vary.

A numerical illustration will make it clearer as to why the *df* used here is $N—1$. If we have 7 scores 4, 6, 7, 8, 10, 12, and 16, the mean of which is 9. We now use this value as estimate of the population mean. The deviations of the scores from the mean 9 are—5,—3,—2, —1,+1,+3, and+7. A mathematical requirement of the mean is that the sum of these deviations is zero. Of the 7 deviations, only 6 i.e. ($N—1$), can be chosen freely (i.e. independently) as the condition that the sum equal to zero restricts the value of the 7th deviate. With this condition we can arbitrarily change any six of the deviates and thereby fix the seventh. We could take the first six deviates as —7, —5, —4, +4, +3, and +1, which would mean that for the sum to equal zero the seventh has to be + 8. Similarly we can try any other changes and if the sum is to remain zero one of the seven deviations is automatically determined. Hence, only 6 i.e. (7—1)'s are free to vary within the restrictions imposed.

When a statistics is used to estimate a parameter, the number of degrees of freedom depends upon the restrictions placed. One *df* is lost for each restriction imposed. Therefore, the number of degrees of freedom will vary from one statistics to another. In estimating and computing population mean (M_{pop}) from the sample mean (M), for example, we lose 1 *df* and so the number of degrees of freedom is $(N-1)$.

By way of illustration, let us determine the .95 and .99 confidence intervals for the population Mean (M_{pop}) of the scores 5, 5, 7,4 and 6 obtained by 5 children in a test of perception. The mean of the scores is $\frac{5+5+7+4+6}{5}$ or 5.40. Using formula (8.4), we compute the standard deviation as:

X	$x=X-M$	x^2
5	-0.4	0.16
5	-0.4	0.16
7	1.6	2.56
4	-1.4	1.96
6	0.6	0.36
	$\Sigma x=0$	$\Sigma x^2=5.20$

$$S=\sqrt{\frac{\Sigma x^2}{N-1}}=\frac{5.20}{5-1}=1.14$$

From formula (8.3), we compute SE_M to be $\frac{1.14}{\sqrt{5}}$ or 0.51

For estimating the M_{pop} from the sample mean of 5.40, we determine the value of t at the selected points using appropriate number of degrees of freedom. The *df* available for determining t are $N-1$ or 4. Entering Table *C* with 4 *df* we read that $t=2.78$ at the .05 point and 4.60 at the .01 point. From the first t, we know that 95 of our 100 sample mean will lie within $\pm 2.78\ SE_M$ or $\pm 2.78 \times 0.51$ of the population mean and that 5 out of 100 fall outside of these limits. The probability (P) is .95, therefore, that our sample mean 5.40 does not miss the M_{pop} by more than $\pm 2.78 \times 0.51$ or ± 1.92 From the second t, we know that 99 per cent of our sample means will lie between M_{pop} and $\pm 4.60\ SE_M$ or $\pm 4.60 \times 0.51$ and that 1 per cent fall beyond these limits. The probability (P) is .99, therefore, that our sample mean of 5.40 does not miss the M_{pop} by more than $\pm 4.60 \times 0.51$ or ± 2.35.

Taking as our limits, $M \pm 2.78 \times SE_M$ we have $5.40+2.78 \times .51$ or 3.98 and 6.82 as the limits of the .95 confidence interval. The probability (P) is .95 that the M_{pop} is not less than 3.98 nor greater than 6.82 Taking the limits $M \pm 4.60\ SE_M$, we have $5.40 \pm 4.60 \times .51$ or

3.05 and 7.7 as the limits of .99 confidence interval, and the probability (P) is .99 that M_{pop} is not less than 3.05 nor greater than 7.75.

In educational research we should avoid to build generalizations on the basis of small samples as it is difficult to be sure that a small sample adequately represents the population from which the sample is drawn. Furthermore, conclusions drawn from small samples are usually unsatisfactory because of great variability from sample to sample.

Significance of the Difference Between Means

Some investigations require the use of a statistical technique to determine whether a true difference exists between population means of two samples. Suppose we wish to determine whether the populations of 13 years old boys and girls differ in abstract reasoning. First, we would draw a large random sample of boys and girls. Next, we would administer a test of abstract reasoning on the samples, compute the means of the two groups, and find the difference between them. Let the mean of the boys be 30 and that of the girls is 28. Then it has to be ascertained if the difference of 2 between the means is large enough to be taken as real and dependable i.e., not due only to sampling errors or chance. We cannot categorically conclude that the mean of boys is large than the mean of girls and the difference of 2 between the two means is significantly large to be taken as dependable. In order to test the significance of an obtained difference, we must first have a standard error of the difference of the two means as it is reasonable to expect that the difference between two means will be subject to sampling errors. Then from the difference between the sample means and its standard error we can determine whether a difference probably exists between the population means.

The Significance of the Difference Between the Independent Means in Large and Small Samples

Means are said to be independent or uncorrelated when computed from samples drawn at random from totally different and uncorrelated groups.

Large Samples. We have already found that the frequency distribution of large sample means drawn from the same population fall into a normal distribution around M_{pop} as their measure of central tendency; and it is reasonable to expect that the frequency distribution of the difference between the means computed from samples drawn from two different populations will also tend to be normal, with a mean of zero and a standard deviation that is called the standard error of the difference means. Such a standard error is denoted by σ_{dM} and can be estimated from the standard errors of the two means, σ_{M1} and σ_{M2}. The formula is

$$\sigma dM = \sqrt{\sigma_{M1}^2 + \sigma_{M2}^2} \qquad (8.5)$$

in which

σM_1=the *SE* of the mean of the first sample
σM_2=the *SE* of the mean of the second sample
N_1=number of cases in first sample
N_2=number of cases in second sample.

By way of illustration, let us apply formula (8.5) to a typical problem. Two groups, one of 95 boys and the other of 64 girls of the tenth grade, were given the same word-building test. The results are summarized in Table 8.1.

TABLE 8.1 Means and Standard Deviations of Two Independent Large Samples

Statistics	*Boys*	*Girls*
N	95	64
Mean (M)	24.5	23.4
Standard deviation (σ)	5.12	4.07

Assuming that our samples are random, it is to be ascertained whether the difference between the means 24.5 and 23.4 is significant.

Using formula (8.5), we compute the standard error of the difference

$$\sigma dM = \sqrt{\frac{(5.12)^2}{95} + \frac{(4.07)^2}{64}}$$
$$= 0.73$$

The obtained difference between the means of the boys and girls is 1.1 ((*i.e.*, 24.5—23.4); and the *SE* of this difference (σdM) is 0.73.

As a first step in determining whether boys and girls actually differ in word-building ability, we shall set up a null hypothesis. This hypothesis asserts that the difference between the population means of boys and girls is zero and that, except for sampling errors, mean differences from sample to sample will all be zero. In accordance with the null hypothesis, we assume a sampling distribution of differences to be normal with the mean at zero, or at M_{pop} (boys)—M_{pop} (girls)=0. The deviation of each sample difference, M_{boys}—M_{girls}, from this central reference point is equal to [M_{boys}—M_{girls}]—[M_{pop}

(boys)—M_{pop} (girls)], or [M_{boys}—M_{girls}]—0. The deviation of each sampled difference given in terms of standard measure would be the deviation divided by the standard error, which gives us a $\overline{Z}$—value in terms of a general formula.

$$\overline{Z}=\frac{|M_1-M_2|}{\sigma dM} \qquad (8.6)$$

Figure 8.3 shows a sampling distribution of $\overline{Z}$ ratios.

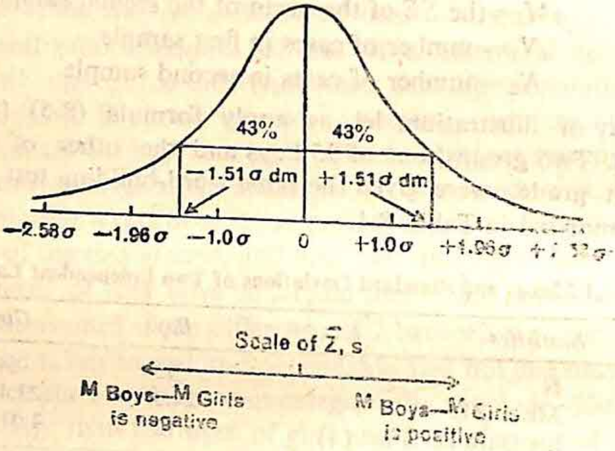


FIG. 8.3. Sampling Distribution of $\overline{Z}$ with a Mean of Zero.

Using formula (8.6)

$$\overline{Z}=\frac{1.10}{0.73}=1.51$$

From the Table *A* we know that 43 per cent × 2 or 86 per cent of the cases in a normal distribution fall between the mean and ±1.51 σdM; and 14 per cent of the cases fall outside. This means that under the null hypothesis we can expect $\overline{Z}$'s as large as or larger than ±1.51 to occur "by chance" 14 times in 100 comparisons of the means of samples of tenth grade boys and girls on the test of word-building.

Levels of Significance

For the sake of convenience, the researchers have .05 and .01 levels of significance as two arbitrary standards for accepting or rejecting a null hypothesis. From Table *A* we read that ± 1.96 σ mark off points along the base line of a normal distribution of the left and right of which lie 5 per cent of the cases. (2.5 per cent at each end of the curve). When a value is 1.96 or $\bar{Z}$ more, we reject a null hypothesis at .05 level of significance. The $\bar{Z}$ value of 1.51 in our problem falls short of 1.96 *i.e.* does not reach the .05 level. Accordingly, we retain the null hypothesis and conclude that tenth grade boys and girls actually do not differ in mean performance on word building test.

Furthermore, from the Table *A* we know that ± 2.58 σ mark off points to the left and right of which lie 1 per cent (0.5 per cent at each end of the curve) of the cases in a normal distribution. If the $\bar{Z}$ value is 2.58 or more, we reject the null hypothesis at .01 level and the probability (*P*) is that not more than once in 100 trials would a difference of this size arise if the true difference $(M_{pop1}—M_{pop2})$ were zero.

Two-tailed and One-tailed Tests of Significance

Suppose a null hypothesis were set up that there was no difference, other than a sampling error difference, between the mean test anxiety score of high school boys and girls, we would be concerned only with a difference, and not in superiority or inferiority in test anxiety of either group. To test this hypothesis we apply two-tailed test as the difference between the obtained means may be as often in one direction (plus) as in the other (minus) from the true difference of zero; and in determining probabilities we take both tails of sampling distribution.

For a large sample two-tailed test, we make use of a normal distribution curve. The 5 per cent area of rejection is divided equally between the upper and lower tails of this curve and we have to go out to ± 1.95 on the base line of the curve to reach the area of rejection.

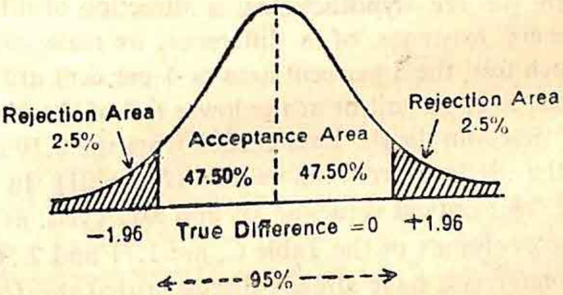


FIG. 8.4. A Two-tailed Test at .05 level (2.5 per cent at each level).

Similarly, we have 0.5 per cent area at each end of the normal curve when 1 per cent area of rejection is to be divided equally between its upper and lower tails and it is necessary to go out to ±2.58 on the base line to reach the area of rejection as shown in Figure 8.5.

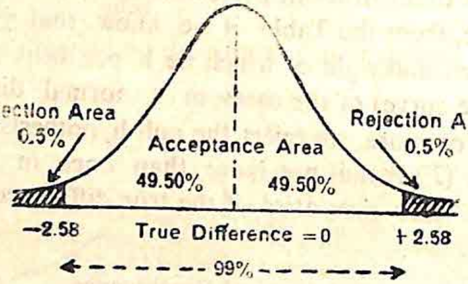


FIG. 8.5. A Two-tailed Test at .01 Level (0.5 per cent at each level).

For a small sample two-tailed test, we make use of *t*-distribution. The *t*-critical values necessary for the rejection of the null hypothesis are determined from the Table *C* of the Appendix with the help of degrees of freedom. For example, in comparing the significance of mean test anxiety score difference between samples, 10 high school boys and 15 high school girls, the number of degrees of freedom is N_1+N_2-2 or $10+15-2$ or 23. From the Table *C* at 23 *df*, the *t* critical values necessary for the rejection of the null hypothesis at .05 and .01 levels are 2.07 and 2.81 respectively.

In the above problem, if we change the null hypothesis as: high school boys do dot have higher test anxiety than high school girls; high school boys do not have lower test anxiety than high school girls; then each of these hypotheses indicates a direction of the difference. When we are hypothesizing a direction of difference, rather than the mere existence of a difference, we make use of one-tailed test. For such test, the 5 per cent area or 1 per cent are a of rejection is either at the upper tail or at the lower tail of the curve. For given degrees of freedom the 0.5 level is read from the 0.10 column ($P/2=.05$) and the .01 level from .02 column ($P/2=.01$). In the one-tailed test, for 23 *df t* critical values at .05 and .011 evels, as read from the 0.10 and 0.02 columns of the Table *C*, are 1.71 and 2.50 respectively.

Small samples. We have already discussed that the frequency distribution of small sample means drawn from the same population forms a *t*-distribution; and it is reasonable to expect that the sampling distribution of the difference between the means computed from small samples drawn from two different populations will also fall under the category of *t*-distribution. The formula provided by Fisher for testing a difference between means computed from uncorrelated samples is

$$t=\frac{|\,M_1-M_2\,|}{\sqrt{\left(\frac{\Sigma x_1{}^2+\Sigma x_2{}^2}{N_1+N_2-2}\right)\left(\frac{N_1+N_2}{N_1N_2}\right\}}} \qquad (8.7)$$

in which

M_1 and M_2 =means of the two samples
$\Sigma x^2{}_1$ and $\Sigma x_2{}^2$=sums of squares of the deviations from the means in the two samples.
N_1 and N_2=numbers of cases in the two samples.

The appropriate *t* critical value for acceptance or rejection of the null hypothesis would be found for N_1+N_2-2 degrees of freedom, using the Table C.

There are some conditions that must be met to justify the use of *t* test for small samples. Guilford (1965, pp. 184-185) presents the undermentioned conditions which should be fulfilled in the use of *t* test.

1. The population distribution should be normal. If the population distribution is seriously skewed, *t*-tests should not be used. For skewed distribution, a number of distribution free tests applicable, same of which are described in non-parametric tests.

2. The two samples under investigation should have the same variances. This condition is known as homogeneity of variances. Whether or not two sample variances are significantly different may be determined by using Bartlett's test of homogeneity which is described later in this chapter in the technique of analysis of variance.

To illustrate the use of formula (8.7), let us test the significance of the difference between mean scores of 13-year old 5 boys and 8 girls on an Interest Test.

$$N_1+N_2--2=5+8-2$$

Using formula (8.7)

$$t=\frac{|\,26-25\,|}{\sqrt{\left\{\frac{104+226}{5+8-2}\right\}\left\{\frac{5+8}{5\times 8}\right\}}}$$

$$=\frac{1}{\sqrt{\left(\frac{330}{11}\right)\left\{\frac{13}{40}\right)}}$$

$$=\frac{1}{\sqrt{9.75}}$$

$$=0.32$$

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

TABLE 8.6 Scores of 5 Boys and 8 Girls on Interest Test

Boys	$(N_1=5)$		*Girls*	$(N_2=8)$	
X_1	x_1	x_1^2	X_2	x_2	x_2^2
20	-6	36	34	9	81
22	-4	16	25	0	0
30	4	16	16	-9	81
32	6	36	30	$+5$	25
26	0	0	22	-3	9
$\Sigma X_1=130$		$\Sigma x_1^2=104$	27	2	4
$M_1=\frac{\Sigma X_1}{N_1}$			20	-5	25
			26	1	1
$=\frac{130}{5}$			$\Sigma X_2=200$		$\Sigma x_2^2=226$
			$M_2=\frac{\Sigma X_2}{N_2}$		
$=26$			$=\frac{200}{8}$		
			$=25$		

This is a two-tailed test, and the t critical values for rejection of null hypothesis in the Table *C* of the Appendix for $(5+8-2)$ or 11 *df.* are 2.20 at .05 and 3.11 at the .01 levels. Since the obtained t value is 0.32, it does not equal or exceed the t critical valve necessary for rejection of the null hypothesis at the .05 level for 11 *af*; the null hypothesis is accepted and we conclude that there is no significant difference in the mean interest scores of 13-years old boys and girls.

The Significance of the Difference between the Dependent Means in Large and Small Samples

Means are said to be dependent or correlated when obtained from the scores of the same test administered to the same sample upon two occasions or when the same test is administered to equivalent samples in which the individuals who make up the groups have been matched person for person, by one or more characteristics.

Suppose that we have administered a test to a group of students and three weeks later have repeated the test.

The formula used for testing the significance of the difference between means obtained in the initial and final testing is

$$t=\frac{|M_1-M_2|}{\sqrt{\sigma M_1^2+\sigma M_2^2-2r_{12}\,\sigma M_1\,\sigma M_2}} \tag{8.8}$$

in which

σM_1=standard error of the initial test mean

σM_2=standard error of the final test mean

r_{12}=correlation between the scores on initial and final testing.

The following example will illustrate the use of formula (8.8). Twelve students of eighth grade of a certain school were administered an achievement test in arithmetic and then all the students were provided remedial instruction in arithmetic. Three weeks later, the test was administered for the second time. To test the hypothesis that the remedial instruction increases the test scores of students, we would test the significance of the difference between the means of the scores for the first and second trials of the test as shown in the Table 8.7.

TABLE 8.7. Scores of Twelve Students of Eighth Grade on the Trial 1 and Trial 2 of an Achievement Test in Arithmetic

Trial 1 (X_1)	*Trial 2* (X_2)
33	45
57	67
50	58
65	79
36	46
55	63
37	47
30	25
21	30
46	56
37	35
45	57
$\Sigma X_1=512$	$\Sigma X_2=608$

Computing from the scores, we have

$M_1=42.66$, $M_2=50.66$

$\sigma_1=12.63$, $\sigma_2=15.74$

$\sigma M_1=3.65$, $\sigma M_2=4.54$

$r_{12}=0.94$

Using formula (8.8)

$$t=\frac{|\,42.66-50.66\,|}{\sqrt{(3.65)^2+(4.54)^2-2\times 0.94\times 3.65\times 4.54}}$$

$$=4.92$$

Since there are 12 students, we have 12 pairs of scores and 12 differences, the *df* becomes 12—1 or 11. To test the hypothesis that remedial instruction in arithmetic increases test score of student, we would make use of one tailed test. In the one tailed-test, for 11 *df* the .05 level is read from .10 column (P/2=.05) of Table *C* to be 1.80 and the .01 level from the .02 column ($P/2=.01$) is 2.72. Since our t of 4.92 is much larger than the .01 level of 2.72, it may be concluded that the gain from Trial 1 to Trial 2 is significant and the hypothesis that the remedial instruction in arithmetic increases the test score is accepted.

A procedure called the Difference Method is also used to test the significance of the difference between the means of test scores obtained from administration of the same test to the same sample upon two occasions. This method is quicker and convenient when the sample is small. The formula used in this method is

$$t=\frac{M_D}{SE_{MD}} \qquad (8.9)$$

in which

M_D= mean of difference

SE_{MD}=standard error of mean of difference

By way of illustration, let us apply difference method to the data of 12 students in Table 8.7

TABLE 8.7

Trial 1	*Trial* 2	*Difference (D)* (2−1)	$d=(D-M_D)$	d^2
33	45	+12	4	16
57	67	+10	2	4
50	58	+ 8	0	0
65	79	+14	6	36
36	46	+10	2	4
55	63	+ 8	0	0
37	47	+10	2	4
30	25	− 5	−13	169
21	30	+ 9	1	1
46	56	+10	2	4
37	35	− 2	−10	100
45	57	+12	4	16
		$\Sigma D=96$		$\Sigma d^2=354$

Mean of difference= $M_D=\frac{\Sigma D}{N}=\frac{96}{12}=8$

$$\text{Standard deviation of difference} = SD_D = \sqrt{\frac{\Sigma d^2}{N-1}}$$

$$= \sqrt{\frac{354}{11}}$$

$$= 5.67$$

$$\text{Standard error of mean of difference} = SE_{MD} = \frac{SD_D}{\sqrt{N}}$$

$$= \frac{5.67}{\sqrt{12}}$$

$$= 1.64$$

Using formula (8.9)

$$t = \frac{8}{1.64}$$

$$= 4.88$$

The result found by using the difference method has come out to be the same as before within the error of computation.

In some experimental situations, we set up equivalent groups using matching technique. One group acts as the experimental group and the other as the control group. The matching is initially done by pairs so that each person in the first group has a match in the second group. In such situations also we make use of formula (8.8) for testing the significance of the difference between the means of the variable under study for the two groups.

The following problem is typical of many in which the equivalent group technique is useful.

Two groups, *A* and *B*, of seventh grade students were matched for initial ability in a test of general science. Group *A* was taught by the lecture method, and Group *B* by the lecture-demonstration method. Data are given in the Table 8.8.

To test the significance of the difference between the mean final scores, 56.35 and 62.40, on the general science test, we make use of two-tailed test. Using formula (8.8).

$$t = \frac{|\,56.35 - 62.40\,|}{\sqrt{\frac{(6.00)^2}{50} + \frac{(7.31)^2}{50} - 2 \times 0.54 \times \frac{6.00}{\sqrt{50}} \times \frac{7.31}{\sqrt{50}}}}$$

$$= 6.61$$

There are 50 pairs, the degrees of freedom to be used in testing this *t* are (50—1) or 49. Entering Table *C* with 49 *df* we find that *t* at .01 level to be 2.68. Since our *t* of 6.61 is much larger than the .01 level *t* of 2.68, we may reject the null hypothesis and conclude that the

TABLE 8.8 Initial and Final Scores of Students in the Control and Experimental Groups on the General Science Test

	Group A (Control)	*Group B Experimental)*
Number of students	50	50
Mean initial score on the general science test	45.30	45.67
SD of initial scores on the general science test	4.26	4.20
Mean final score on the general science test	56.35	62.40
SD of final scores on the general science test	6.00	7.31
Correlation between final scores on the general science test	0.54	

mean of group *B* (experimental) is significantly higher than the mean of group *A* (control).

At times it is impracticable or impossible to set up groups in which subjects have been matched person for person. In such situations it is desirable to match the subjects in terms of mean and standard deviation of the matching variable.

Suppose that *X* is the variable under study, and *Y* is the variable in terms of which our groups have been matched for mean and standard deviation. Then if r_{xy} is the correlation between *X* and *Y* in the population from which our samples have been drawn, the formula for testing the significance of difference between means is

$$t = \frac{| M_{x1} - M_{x2} |}{\sqrt{\sigma^2 M_{x1} + \sigma^2 M_{x2} \, (1 - r^2_{xy})}} \tag{8.10}$$

in which

σM_{x1} = standard error of the mean of the scores in *X* variable for the first group

σM_{x2} = standard error of the mean of the scores in *X* variable for the second group

r_{xy} = Correlation between the variable under study and the variable in terms of which the groups have been matched for mean and SD.

To illustrate the use of formula (8.10), suppose two groups, *A* and *B*, of high school students were matched for mean and SD upon a group test of general intelligence for their comparison in comprehension of humour. The groups were administered a test in comprehension of homour and the data are given in Table 8.9.

TABLE 8.9. Mean and Standard Deviation of Scores on Comprehension in Humour Test of Matched Groups

	Group A	*Group B*
Mean on comprehension in humour test	45.20	52.27
SD on comprehension in humour test	10.00	13.00
Number of students	60	70

The correlation between the scores in group test of general intelligence and comprehension of humour test for the high school students was found to be 0.50.

To test the significance of the difference between the means 45.20 and 52.27 for the two groups, we make use of two-tailed test. Using formula (8 10)

$$t=\frac{|\,45.20-52.27\,|}{\sqrt{\left(\frac{(10)^2}{60}+\frac{(13)^2}{70}\right)\left(1-(.50)^2\right)}}$$

$$=4.04$$

The degrees of freedom to be used in testing this t are $(N_1-1)+(N_2-1)-1$, or $(60-1)+(70-1)-1$ or 127. Entering Table *C* with 127 *df*, we find that our t of 4.04 is larger than the entry of 2.62 at the .01 level; and therefore, the null hypothesis that the two groups do not differ in comprehension of humour is rejected. Since the mean of group *B* is larger than that of group *A*, we may conclude that group *B* has performed better in the test of humour comprehension than group *A*.

Analysis of Variance

We use $\bar{z}$ and t tests to determine whether there is any singnificant difference between the means of two random samples. Suppose we have six random samples and we want to determine whether there are any significant differences among their means. For this we would have to use $\frac{6(6-1)}{2}=15$ t tests to determine the significance of the difference between the six means by taking two means at a time. This procedure is cumbersome and time consuming. The technique of analysis of variance would make it possible to determine if any of the two of the six means differ significantly from each other by a single test, called F test, rather than 15 t tests. The F test enables us to determine whether the sample means differ from one another (between group variance) to a greater extent than the test scores differ from their own sample means (within groups variance) using the ratio

$$F = \frac{\text{Variance between groups}}{\text{Variance within groups}} \tag{8.11}$$

The significance of F ratio is determined from the Table D of the Appendix. This table indicates the F critical values necessary to reject the null hypothesis at selected levels of significance.

Basic Assumptions of Analysis of Variance

Analysis of variance has certain basic assumptions underlying it. Johnson (1961) presents certain assumptions which should be fulfilled in the use of this technique.

1. The population distribution should be normal. This assumption, however, is not especially important. Eden and Yates showed that even with a population departing considerably from normality, the effectiveness of the normal distribution still held. Besides the findings of Eden and Yates, the study of Norton cited by Guilford (1965, pp. 300-301) also points out that F is rather insensitive to variations in the shape of population distribution.

2. All the groups of a certain criterion or of the combination of more than one criterion should be randomly chosen from the sub-population having the same criterion or having the same combination of more than one criterion. For instance, if we wish to select two groups in a school population, one of the third grade and the other of the fourth grade, we must choose randomly from the respective sub-populations. This assumption is the key stone of the analysis of variance technique. Failure to fulfil this assumption gives biased results.

3. The sub-groups under investigation should have the same variability. We should test this assumption before we run through the analysis of variance. Otherwise, a false interpretation of the results may follow. This assumption is tested either by applying Bartlett's test of homogeneity or by applying Hartley's test.

Analysis of Variance for a Randomized Group Design (One-way Analysis of Variance)

In a learning experiment, 28 subjects were randomly assigned to each of the three groups. Each group performed the same task but under slightly different experimental treatments. The outcomes of the experiment are given in the Table 8.10.

TABLE 8.10. Data of the Groups Obtained Under Three Experimental Treatments

Treatment 1 (X_1)	*Treatment 2* (X_2)	*Treatment 3* (X_3)	
15	13	20	
17	18	28	
16	18	18	
11	13	11	
19	16	16	
16	14	15	
15	20	16	
18	21	19	
18	18	11	
18	13	9	
14	14	14	
18	19	15	
11	14	14	
11	22	14	
14	16	12	
11	16	19	
17	12	20	
20	21	24	
18	16	14	
12	13	16	
12	17	21	
21	20	20	
11	12	9	
17	18	18	
16	15	14	
12	15	17	
12	13	16	
9	15	8	
ΣX 419	452	448	1319
ΣX^2 6561	7532	7726	21819

Te test the hypothesis that the groups do not differ in mean performance, let us apply the technique of analysis of variance.

The basic assumptions underlying this technique are tested as under:

1. *Assumption of normality*. In the light of the findings of Eden and Yates (Johnson, 1961), and Norton (Guilford, 1965, pp. 300-301) the assumption of normality may not be considered important for the data of the experiment.

2. *Assumption of randomness.* The requirement of randomness has been amply fulfilled in this experiment as the subjects were randomly assigned to three experimental groups from the same pool sample.

3. *Assumption of homogeneity of variance.* This assumption may be tested by applying Bartlett's test of homogeneity of variance to the data of the three groups in the following manner.

TABLE 8.11 Bartlett's Test of Homogeneity of Variance for the Data of the Groups Obtained Under Three Expeeimental Treatments

Treatment *(K)*	*df* *(N—1)*	ΣXK^2	SK^2	*log* SK^2
1	27	290.96	10.78	1.0326
2	27	235.43	8.72	0.9405
3	27	558.00	20.67	1.3153
			$\Sigma SK^2=40.17$	Σ log $SK^2=3.2884$

in which

$$\Sigma XK^2=\Sigma X^2-\frac{(\Sigma X)^2}{N} \tag{8.12}$$

$$SK^2=\frac{\Sigma XK^2}{N-1} \tag{8.13}$$

Computations

1. $\log\frac{\Sigma SK^2}{K}=\log\frac{40.17}{3}=\log 13.39=1.1268$

2. $K\log\frac{\Sigma SK^2}{K}=3\times 1.1268=3.3804$

3. $\text{Difference}=K\log\frac{\Sigma SK^2}{K}-\Sigma\log SK^2$

$$=3.3804-3.2924=0.092$$

4. χ^2 (Chi-square)$=2.3026\times(N-1)$ (Difference) (8.14)
in which 2.3026 is a constant term

$$=2.3026\times 27\times 0.092$$
$$=5.7196$$

5. $\text{Correction (C)}=1+\frac{K+1}{3K(N-1)}$ (8.15)

$$=1+\frac{3+1}{3\times 3\times 27}$$

$$=1+\frac{4}{243}=1.016$$

6. Corrected χ^2(Chi-square)$= \dfrac{\chi^2}{\text{correction}}$ (8.16)

$$= \frac{5.7196}{1.016}$$

$$=5.6296$$

7. $df=K-1=3-1=2$

The value of χ^2 needed for significance for 2 df at .05 level is 5.991 (Table E of the Appendix). Since the obtained value 5.6296 of χ^2 is less than the needed value, the variance within the cells may be treated as homogeneous.

After testing the data for the basic assumptions we may proceed with the computations for the analysis of variance.

1. *Correction*$=\dfrac{(\Sigma X)^2}{N}=\dfrac{(1319)^2}{84}=20711.44$

2. *Total sum of squares* (*Total SS*)

$=\Sigma X^2-\text{correction}$

$=15^2+17^2+\dots\dots\dots+16^2+8^2-20711.44$

$=1107.56$

3. *Sum of squares between means of treatment 1, 2 and 3* (*SS between Means*)

$$=\frac{(\Sigma X_1)^2}{N_1}+\frac{(\Sigma X_2)^2}{N_2}+\frac{(\Sigma X_3)^2}{N_3}-\text{correction}$$

$$=\frac{(419)^2}{28}+\frac{(452)^2}{28}+\frac{(448)^2}{28}-20711.44$$

$$=\frac{175561+204304+200704}{28}-20711.44$$

$$=20734.61-20711.44$$

$$=23.17$$

4. *Sum of squares within treatments 1, 2 and 3* (*SS within Means*)

$=$Total $SS-SS$ Between Means

$=1107.56-23.17=1084.39$

5. *Calculation of variances from each SS and analysis of the total variance into its components*. Each SS becomes a variance when divided by the degrees of freedom (df) allotted to it. There are 84 scores in all in the Table 8.10, and hence there are $(N-1)$ or 83 df in all. These 83 df are allocated in the following way:

If N=number of scores in all and K=number of treatments or groups, we have df for total $SS=N-1=84-1=83$; df for within treatments $=N-K=84-3=81$; and df for between the means of treatments$=K-1=3-1=2$

The variance among means of treatments is 23.17/2 or 11.585; and the variance within means is 1084.39/81 or 13.387.

TABLE 8.12. **Summary: Analysis of Variance**

Source of variation	*d.f*	*Sum of squares (SS)*	*Mean square (variance).*
Between the means of treatments	2	23.17	11.585
Within treatments	81	1084.39	13.387
Total	83	1107.56	

Using formula (8.11)

$$F=\frac{11.585}{13.387}=0.865$$

In the present problem the null hypothesis asserts that three sets of scores are in reality the scores of three random samples drawn from the same normally distributed population, and that the means of the treatments 1, 2 and 3 will differ only through the fluctuations of sampling. To test this hypothesis we divided the "between means" variance by the "within treatments" variance and compared the resulting variance ratio, called *F*, within the *F* values in Table *D* of the Appendix. The *F* in our problem is 0.865 and the *df* are 2 for the numerator (*df* 1) and 81 for the denominator (*df* 2). Entering Table *D* we read from column 2 and row 81 (interpolate between rows 80 and 90) that an *F* of 3.11 is significant at .05 level and *F* of 4.88 is significant at the .01 level. Since the obtained *F* value of 0.865 is less than the table values, we reject the null hypothesis, and, therefore, conclude that the means of three groups do not differ significantly.

Analysis of Variance for Factorial Design (Three-way Analysis of Variance)

In many experimental studies, we are concerned with the effect of two or more independent variables, usually called *factors*, on a dependent variable. The number of ways in which a factor is varied is called the number of levels of the factor. A factor that is varied in two ways is said to have two levels and a factor that is varied in three ways is said to have three levels. With two or more factors each with two or more levels, a treatment in some experiments consists of a combination of one level for each factor. Edwards (1971, p. 200) says :

When the treatment consists of all possible different combinations of one level from each factor, and we have an equal number of observations for each treatment, the experiment is described as a *complete factorial experiment with equal* replications.

As an illustration let us suppose that the dependent variable is a measure of retention of verbal material. One independent variable (factor) may be the number of times the material is presented; the second independent variable (factor) may be the mode of presentation of the material; and the third independent variable (factor) may be time of testing. Suppose the first factor (A) is varied in two ways: (*i*) by presenting the material once (A_1) and (*ii*) by presenting the material twice (A_2); the second factor (B) is varied in three ways: (*i*) in one case, the subjects reads the material themselves (B_1: visual mode of presentation); (*ii*) in the second case, the experimenter reads the material to the subjects (B_2; auditory mode of presentation); and (*iii*) in the third case, first the experimenter reads the material to the subjects and then the subjects are asked to read the material themselves (B_3: auditory cum visual mode of presentation); the third factor (C) is also varied in three ways: (*i*) the administration of the test immediately after the experiment (C_1: immediate retention); (*ii*) the administration of the test after two weeks after the experiment (C_2; Two weeks retention); and (*iii*) the administration of the test after six weeks after the experiment (C_3: six weeks retention).

A given treatment can be obtained by selecting one level from each of the three factors. For example, one treatment will be $A_1B_1C_1$ and will represent a treatment consisting of one presentation, using the visual mode, and an immediate test. The total number of different treatments will be $2 \times 3 \times 3 = 18$, shown in Table 8.13.

TABLE 8.13. The Eighteen Treatment Combinations of $2 \times 3 \times 3$ Factorial Experiment

Treatment	*Number*	*Mode*	*Time*
$A_1\,B_1\,C_1$	One	Visual	Immediate
$A_1\,B_1\,C_2$	One	Visual	Two weeks
$A_1\,B_1\,C_3$	One	Visual	Six weeks
$A_1\,B_2\,C_1$	One	Auditory	Immediate
$A_1\,B_2\,C_2$	One	Auditory	Two weeks
$A_1\,B_2\,C_3$	One	Auditory	Six weeks
$A_1\,B_3\,C_1$	One	Visual cum auditory	Immediate
$A_1\,B_3\,C_2$	One	Visual cum auditory	Two weeks

Treatment	*Number*	*Mode*	*Time*
$A_1 B_3 C_3$	One	Visual cum auditory	Six weeks
$A_2 B_1 C_1$	Two	Visual	Immediate
$A_2 B_1 C_2$	Two	Visual	Two weeks
$A_2 B_1 C_3$	Two	Visual	Six weeks
$A_2 B_2 C_1$	Two	Auditory	Immediate
$A_2 B_2 C_2$	Two	Auditory	Two weeks
$A_2 B_2 C_3$	Two	Auditory	Six weeks
$A_2 B_3 C_1$	Two	Visual cum auditory	Immediate week
$A_2 B_3 C_2$	Two	Visual cum auditory	Two weeks
$A_2 B_3 C_3$	Two	Visual cum auditory	Six weeks

Suppose the researcher randomly assigns 20 subjects to each group for each treatment and obtains data on the dependent variable. The outcomes of this hypothetical experiment are given in Table 8.14.

Assumption of Homogeneity of Variance

This assumption may be tested by applying Bartlett's test of homogeneity to the data given in the Table 8.14. Using formulae (8.12) and (8.13) we obtain the values for the Table 8.15.

TABLE 8.15. Bartlett's Test of Homogeneity of Variance for the Data of Groups Obtained Under Eighteen Experimental Treatments

Treatment *(K)*	*df* *(N—1)*	ΣXK^2	SK^2	*log* SK^2
$A_1 B_1 C_1$	19	184.95	9.73	0.9981
$A_1 B_1 C_2$	19	170.80	8.99	0.9583
$A_1 B_1 C_3$	19	292.00	15.37	1.1866
$A_1 B_2 C_1$	19	183.20	9.64	0.9841
$A_1 B_2 C_2$	19	142.95	7.52	0.8762
$A_1 B_2 C_3$	19	131.20	6.91	0.8395
$A_1 B_3 C_1$	19	101.20	5.33	0.7267
$A_1 B_3 C_2$	19	156.80	8.25	0.9165
$A_1 B_3 C_3$	19	145.75	7.67	0.8848
$A_2 B_1 C_1$	19	230.55	12.13	1.0838
$A_2 B_1 C_2$	19	112.95	5.94	0.7738
$A_2 B_1 C_3$	19	203.20	10.69	1.0289
$A_2 B_2 C_1$	19	352.55	18.56	1.2685
$A_2 B_2 C_2$	19	152.95	8.05	0.9058
$A_2 B_2 C_3$	19	230.95	12.16	1.0847
$A_2 B_3 C_1$	19	187.00	9.84	0.9930
$A_2 B_3 C_2$	19	136.20	7.17	0.8554
$A_2 B_3 C_3$	19	156.00	8.21	0.9143
			$\Sigma SK^2=172.16$	$\Sigma \log SK^2=17.2645$

TABLE 8.14. Data of the Groups Obtained Under Different Experimental Treatments

	A₁									A₂									
	B₁			B₂			B₃			B₁			B₂			B₃			
	C₁	C₂	C₃	C₁	C₂	C₃	C₁	C₂	C₃	C₁	C₂	C₃	C₁	C₂	C₃	C₁	C₂	C₃	
	19	18	17	20	22	19	20	15	18	22	19	11	10	23	15	21	17	16	
	17	21	16	20	22	19	17	18	22	22	14	15	12	23	15	22	19	14	
	14	16	11	23	19	15	22	17	18	14	14	17	13	13	22	24	18	22	
	18	16	21	17	18	19	16	18	19	17	15	15	13	15	22	24	20	21	
	23	19	13	18	18	23	18	16	15	16	17	25	9	17	21	16	17	14	
	17	9	25	20	19	16	19	19	17	20	18	17	14	18	18	20	19	19	
	18	18	19	19	21	19	18	17	17	16	14	17	16	19	13	21	18	19	
	15	19	20	22	17	18	14	19	14	22	18	21	11	14	12	19	20	18	
	21	19	15	22	22	19	20	19	18	18	16	20	9	20	14	18	19	16	
	16	14	18	17	17	16	18	11	16	21	15	13	11	18	23	14	22	20	
	17	17	22	11	17	17	18	13	20	20	14	17	8	15	19	18	18	14	
	14	15	20	20	16	20	18	17	22	19	19	16	12	17	23	20	20	19	
	19	20	17	18	14	17	13	14	17	14	14	19	19	16	15	14	19	18	
	20	19	19	12	22	18	20	23	18	13	20	12	17	19	15	16	21	21	
	19	15	11	18	20	14	17	15	20	17	19	16	23	20	18	15	19	13	
	23	23	18	17	19	13	16	13	13	22	21	19	21	17	20	22	22	16	
	13	18	19	16	16	19	19	14	20	13	14	18	18	21	17	16	22	20	
	18	16	15	18	17	15	14	14	12	16	18	16	14	21	22	16	21	23	
	19	16	14	21	13	12	19	16	21	12	17	20	19	15	16	16	24	18	
	11	20	10	15	22	16	20	20	18	13	13	20	18	18	19	18	11	19	
ΣX	351	348	340	364	371	344	356	328	355	347	329	344	287	359	359	370	386	360	6298
ΣX^2	6345	6226	6072	6808	7025	6048	6438	5536	6447	6251	5525	6120	4471	6597	6675	7032	7586	6636	113838

Computations

1. $\log\frac{\Sigma SK^2}{K}=\log\frac{172.16}{18}=\log 9.56=0.9805$

2. $K\log\frac{\Sigma SK^2}{K}=18\times 0.9805=17.6490$

3. $\text{Difference}=K\log\frac{\Sigma SK^2}{K}-\Sigma\log SK^2$

$$=17.6490-17.2645$$
$$=0.3845$$

4. Using formula (8.14)

$$\chi^2\ (\text{Chi-square})=(2.3026)\ (N-1)\ (\text{difference})$$
$$=(2.3026)\ (19)\ (0.3845)$$
$$=16.8216$$

5. Using formula (8.15)

$$\text{Correction (C)}=1+\frac{K+1}{3K\ (N-1)}$$
$$=1+\frac{18+1}{3\times 18\times 19}$$
$$=1+\frac{1}{54}=1.0185$$

6. Using formula (8·16)

$$\text{Corrected } \chi^2\ (\text{chi-square})=\frac{\chi^2}{\text{correction (C)}}$$
$$=\frac{16.8216}{1.0185}$$
$$=16.52$$

7. $df=K-1=18-1=17$

The value of χ^2 needed for significance for 17 *df* at .05 level is 27.587. Since the obtained value 16.52 of χ^2 is less than the needed value, the variance within the cells may be treated as homogeneous.

After testing the data for the homogeneity of variance we may proceed with the computations for the analysis of variance.

1. *Correction*$=\frac{(\Sigma X)^2}{N}=\frac{(6298)^2}{360}=110180.01$

2. *Total sum of squares (Total SS)*

$$=\Sigma X^2-\text{correction}$$
$$=[(19)^2+(17)^2+(14)^2+\ldots\ldots+(23)^2+(18)^2+(19)^2]-110180.01$$
$$=3657.99$$

3. *Between and within sum of squares.* For calculating these values the following Table 8.16 is set up. It shows the sum of the twenty scores under each of 18 treatments.

TABLE 8.16. Sums of Twenty Observations Under the Different Treatments

		C_1	C_2	C_3	Σ
	B_1	351	348	340	1039
A_1	B_2	364	371	344	1079
	B_3	356	328	355	1039
	B_1	347	329	344	1020
A_2	B_2	287	359	359	1005
	B_3	370	386	360	1116
	Σ	2075	2121	2102	6298

4. *Sum of squares between treatments (SS between Means)*

$$=\left[\frac{\Sigma X_1{}^2}{N_1}+\frac{\Sigma X^2{}_2}{N_2}+\ldots\ldots+\frac{\Sigma X_{18}{}^2}{N_{18}}\right]-\text{correction}$$
$$=\left[\frac{(351)^2}{20}+\frac{(348)^2}{20}+\ldots\ldots+\frac{(360)^2}{20}\right]-110180.01$$
$$=110566.80-110180.01$$
$$=386.79$$

5. *Sum of squares within treatments (SS within Means)*

$$=\text{Total } SS-SS \text{ between Means}$$
$$=3657.99-386.79$$
$$=3271.20$$

6. *Partitioning the treatment sum of squares.* The sum of squares between treatments which is based upon (20 − 1) or 19 *df* can be divided into different components i.e., main effects and interaction.

TABLE 8.17. Two-way Table for Variables B and C Summed over the Levels of A

	C_1	C_2	C_3	Σ
B_1	698	677	684	2059
B_2	651	730	703	2084
B_3	726	714	715	2155
Σ	2075	2121	2102	6298

From the data given in Table 8.16 we set up the table for the level of B and C summed over the levels of A.

Main Effects for B and C

Each cell entry in Table 8.17 will be the sum of $2 \times 20 = 40$ observations because the variable A has got two levels: A_1 and A_2. The totals of rows 2059. 2084 and 2155 are the sums of three levels of B_1 B_2 and B_3 respectively and these sums are based upon $3 \times 40 = 120$ observations. Similarly column totals 2075, 2121 and 2102 are the sums of $3 \times 40 = 120$ observations.

Sum of squares for B (SS_B)

$$= \frac{1}{120}\left[(2059)^2 + 2084)^2 + (2155)^2 \right] - \text{correction}$$
$$= 110221.35 - 110180.01$$
$$= 41.34$$

Sum of squares for C (SS_C)

$$= \frac{1}{120}\left[(2075)^2 + (2121)^2 + (2102)^2 \right] - \text{correction}$$
$$= 110188.92 - 110180.01$$
$$= 8.01$$

Double interaction $B \times C$

The sum of squares between the nine sums entered in the cells of the Table 8.17 is based on 40 observations.

Sum of squares between cells (SS cells)

$$= \frac{1}{40}\left[(698)^2 + (677)^2 + \ldots\ldots + (715)^2 \right] - \text{correction}$$
$$= 110309.90 - 110180.01$$
$$= 129.89$$

The $B \times C$ interaction sum of squares may be obtained by subtracting the sum of squares for B and **C** from the sum of squares between the cells.

Sum of squares for $B \times C$

$$= SS\ cells - SS_B - SS_C$$
$$= 129.89 - 41.34 - 8.91$$
$$= 79.64$$

Again from the data of Table 8.16 we set up another two-way table for variables A and B summed over levels of C.

TABLE 8.18. Two-way Table for Variables A and B Summed Over the Levels of C

	B_1	B_2	B_3	Σ
A_1	1039	1079	1039	3157
A_2	1020	1005	1116	3141
Σ	2059	2084	2155	6298

Main Effect for A

Each cell entry in the Table 8.18 is based upon 60 observations and the sums 3157 and 3141 are the sums for the levels of A_1 and A_2 respectively which are based upon $3 \times 60 = 180$ observations.

Sum of squares for A (SSA)

$$= \frac{1}{180}\left[(3157)^2+(3141)^2\right] - \text{correction}$$

$$= 110180.72 - 110180.01$$

$$= 0.71$$

Double interaction A × B

The sum of squares for *B* has already been calculated from Table 8.17, and is equal to 41.34. Now to calculate the $A \times B$ interaction sum of squares, we first find out the sum squares between cells of the Table 8.18.

Sum of sqares between cells (SS cells)

$$\frac{1}{60}\left[(1039)^2+(1079)^2+(1039)^2+(1020)^2+(1005)^3+1116)^2\right] - \text{correction}$$

$$= 110319.40 - 110180.01$$

$$= 139.39$$

The $A \times B$ interaction sum of squares may be calculated by subtracting sum of squares for *A* and *B* from the sum of squares between the cells.

Sum of the squares between A × B

$$= 139.39 - 0.71 - 41.34$$

$$= 97.34$$

Double interaction A × C

Again we set up another two-way table to obtained the sum of squares for $A \times C$ interaction.

TABLE 8.19. Two-way Table for Variables A and C Summed Over the Levels of B

	C_1	C_2	C_3	Σ
A_1	1071	1047	1039	3157
A_2	1004	1074	1063	3141
Σ	2075	2121	2102	6298

Sum of squares between (SS cells)

$$=\frac{1}{60}\Big[(1071)^2+(1047)^2+(1039)^2+1004)^2$$
$$+(1074)^2+(1063)^2\Big]-\text{correction}$$
$$=110237.20-110180.01$$
$$=57.19$$

The sum of squares for the $A \times C$ interaction may be calculated by subtracting the sum of squares for A and C from the sum of squares of the cells.

Sum of squares between $C \times A$

$$=57.19-0.71-8.91$$
$$=47.57$$

Triple Interaction A × B × C

We know that the total of the sums of squares for A, B, C, $A \times B$, $B \times C$, $C \times A$ and $A \times B \times C$ must be equal to the sum of squares between the treatments.

$$SS_A+SS_B+SS_C+SS_{A\times B}+SS_{B\times C}+SS_{CXA}$$
$$=0.71+41.34+8.91+97.34+79.64+47.57$$
$$=275.51$$

The sum of the squares between $A \times B \times C$ can be obtained by subtracting the sum 275.51 from 386.79, the sum of the squares between treatments.

Sum of the squares between $A \times B \times C$

$$=386.79—275.51$$
$$=111.28$$

7. *Degrees of freedom (df)*. The analysis of variance has resulted in the partitioning of the total sum of squares and degrees of freedom into two parts. One part is associated with differences among eighteen treatment means and is based on $K-1$ or $18-1=17$ *df*. The other part is associated with the variation within each of the treatment

groups and has (*K*) ($N-1$)=(18) (20−1)=342 *df*. The *A* sum of squares is based on a comparison of the sums for one and two presentations of the material. This sum of squares has (2−1) or 1 *df*. The *B* sum of squares is based on a comparison of the sums for visual, auditory and visual cum auditory modes of presentation. This sum of squares has 3−1 or 2 *df*. The *C* sum of squares is based on the sums for the immediate, two weeks and six weeks tests. This sum of squares also has 3−1 or 2 *df*.

A general rule for determining the degrees of freedom for any interaction sum of squares is to multiply the degrees of freedom associated with the factors for which the interaction is being computed. In the present problem the degrees of freedom for the interactions $A\times B$ and $A\times C$ will be (2−1) (3−1)=2 *df*. Similarly the degrees of freedom for $B\times C$ will be (3−1) (3−1)=4 *df*. The degrees of freedom associated with $A\times B\times C$ interaction is the product of the degrees of freedom associated with the factors involved in the interaction. Because we have 1 *df* for *A*, 2 *df* for *B* and 2 *df* for *C*, the *df* for $A\times B\times C$ interaction sum of squares will be $1\times 2\times 2=4$ *df*.

8. The summary of the complete analysis of variance is presented in the Table 8.20.

The values of *F* that have been entered in the Table 8.20 are obtained by dividing each of the mean squares that is to be tested for significance by the error mean square, that is, within treatment mean square. Thus the value of *F* for *A* is based on 1 and 342 *df*. Each value of *F* for *B*, *C*, $A\times B$ and $C\times A$ is based on 2 and 342 *df*. Similarly each value of *F* for $B\times C$ and $A\times B\times C$ is based on 4 and 342 *df*.

TABLE 8.20. Summary: Analysis of Variance

Source of variation		*Sum of squares*	*df*	*Mean squares*	*F*
A	Number	0.71	1	0.71	0.07
B	Mode	41.34	2	20.57	2.16
C	Time	8.91	2	4.555	0.47
$A\times B$	Number×Mode	97.34	2	48.67	5.09
$B\times C$	Mode×Time	79.64	4	19.91	2.08
$C\times A$	Number×Time	47.57	2	23.785	2.49
$A\times B\times C$	Number×Mode× Time	111.28	4	27.82	2.91
Error: Within treatments		3271.20	342	9.57	
Total		3657.99	359		

9. *Summary of conclusions: main effects.* Entering Table *D*, we find that the main effects for A, B and C are not significant at .05 level of significance. The non-significance of A mean square tells us that the means for A_1 and A_2, averaged over levels of B and C, do not differ significantly. Similarly we can interpret the non-significance of B and C main effects.

Interaction. The $A \times B$ interaction is significant. Therefore, the A effect, that is, the difference between A_1 and A_2, or between one and two presentations, is dependent on the particular mode of presentation employed. The interactions $B \times C$ and $C \times A$ are not significant.

The significant $A \times B \times C$ interaction means: (1) that the $A \times$ B interactions for the separate levels of C are not of the same form; (2) that the $C \times A$ interactions for the separate levels of B are not of the same form; (3) and that the $B \times C$ interactions for the separate levels for A are not of the same form.

Analysis of co-variance. Analysis of co-variance, represents an extension of analysis of variance that tests the significance of difference between means of final experimental data by taking into account the correlation between the dependent variable and one or more co-variates or pertinent control variables, and by adjusting initial mean differences in the groups. It is useful to experimental psychologist, especially in the field of memory and learning, when for various reasons it is not possible or quite difficult to equate experimental groups on some pertinent variable at the start of the experiment.

In an experiment, suppose the dependent variable was the performance of two groups of students on a criterion test; one group was taught through programmed learning (experimental group) and the other through conventional method (control group). The two groups were framed by randomly assigning 70 subjects to each of the groups and the subjects were equated on the pre-test scores on the criterion test. Since intelligence was also considered to be a significant factor related to the scores on the criterion test, the scores on the intelligence test were also taken into account so as to provide a statistical control for testing the significance of the difference between the means of criterion test scores of the two groups at the end of the experiment.

Let scores of the subjects of the experimental and control groups on intelligence test be denoted by X_1 and X_2 and those on the criterion test (post test) as Y_1 and Y_2 respectively. The total number of subjects in experimental and control groups be denoted by N_1 and N_2 respectively. Moreover, the sum of the scores and the sum of squares of the scores on intelligence and criterion tests obtained by the subjects of the two groups be denoted by the following symbols:

ΣX_1 = the sum of the scores on intelligence test obtained by the subjects of the experimental group.

ΣX_2 = the sum of the scores on intelligence test obtained by the subject of the control group.

$\Sigma X = \Sigma X_1 + \Sigma X_2$

ΣX_1^2 = the sum of the squares of the scores on intelligence test obtained by the subjects of the experimental group.

ΣX_2^2 = the sum of the squares of the scores on intelligence test obtained by the subject of the control group.

$\Sigma X^2 = \Sigma X_1^2 + \Sigma X_2^2$

ΣY_1 = the sum of the scores on criterion test obtained by subjects of the experimental group.

ΣY_2 = the sum of the scores on criterion test obtained by the subjects of the control group.

$\Sigma Y = \Sigma Y_1 + \Sigma Y_2$

ΣY_1^2 = the sum of the squares of the scores on criterion test obtained by the subjects of the experimental group.

ΣY_2^2 = the sum of the squares of the scores on criterion test obtained by the subjects of the control group.

$\Sigma Y^2 = \Sigma Y_1^2 + \Sigma Y_2^2$

$\Sigma X_1 Y_1$ = the sum of the products of the scores on intelligence and criterion test obtained by the subjects of the experimental group.

$\Sigma X_2 Y_2$ = the sum of the products of the scores on intelligence test and criterion test obtained by the subjects of control group.

$\Sigma XY = \Sigma X_1 Y_1 + \Sigma X_2 Y_2$

N = the total number of subjects in the experimental and control groups.
$= N_1 + N_2$

The outcomes of the experiment are given in the Table 8.21.

$\Sigma X = \Sigma X_1 + \Sigma X_2$
$= 3188 + 3283$
$= 6471$

$\Sigma X^2 = \Sigma X_1^2 + \Sigma X_2^2$
$= 151488 + 161359$
$= 312847$

$\Sigma XY = \Sigma X_1 Y_1 + \Sigma X_2 Y_2$
$= 257968 + 256161$
$= 514129$

$\Sigma Y = \Sigma Y_1 = \Sigma Y_2$
$= 5608 + 5421$
$= 11029$

$\Sigma Y^2 = \Sigma Y_1^2 + \Sigma Y_2^2$
$= 451038 + 422299$
$= 873337$

TABLE 8.21. Data of the Experimental and Control Groups Obtained under Two-experimental Treatments

	Experimental group			*Control group*	
Subject	*Score on intelligence test* (X_1)	*Score on criterion test* (Y_1)	*Subject*	*Score on intelligence test* (X_2)	*Score on criterion test* (Y_2)
1	50	84	1	37	70
2	35	80	2	47	81
3	28	76	3	46	70
4	43	80	4	58	84
5	47	83	5	39	80
6	41	81	6	42	76
7	29	70	7	52	81
8	46	82	8	30	65
9	29	68	9	44	70
10	28	76	10	32	72
11	43	81	11	50	82
12	46	82	12	49	81
13	35	78	13	52	70
14	38	81	14	57	80
15	39	81	15	40	69
16	49	80	16	24	80
17	41	81	17	43	80
18	46	81	18	38	80
19	43	80	19	46	83
20	45	82	20	40	78
21	59	85	21	54	81
22	46	82	22	52	75
23	43	81	23	48	80
24	38	70	24	53	68
25	49	83	25	48	81
26	59	80	26	46	70
27	53	84	27	44	80
28	52	84	28	40	72
29	48	83	29	60	85
30	56	84	30	47	81
31	42	81	31	48	80
32	33	68	32	70	62
33	54	84	33	52	80
34	58	85	34	27	60
35	63	85	35	49	81
36	58	84	36	49	77
37	57	84	37	50	72
38	60	86	38	51	81
39	23	70	39	42	72
40	53	80	40	24	62

41	63	86	41	68	84
42	48	83	42	46	80
43	44	81	43	47	78
44	40	80	44	60	84
45	46	70	45	61	75
46	55	84	46	59	81
47	26	75	47	52	70
48	48	83	48	53	80
49	51	81	49	27	70
50	42	80	50	43	82
51	24	63	51	50	77
52	53	74	52	38	78
53	44	80	53	39	80
54	51	83	54	61	86
55	52	84	55	60	85
56	47	82	56	49	84
57	48	82	57	51	80
58	50	83	58	63	81
59	47	82	59	19	70
60	56	85	60	51	79
61	39	70	61	47	82
62	46	82	62	38	80
63	62	86	63	40	77
64	37	80	64	49	80
65	46	81	65	28	78
66	53	84	66	50	80
67	39	80	67	56	84
68	46	81	68	59	86
69	49	83	69	52	79
70	31	70	70	50	79

$M_{X_1}=45.54$, $M_{X_2}=46.90$, $GM_X=46.22$, $M_{Y_1}=80.11$
$M_{Y_2}=77.44$

Basic Assumptions of Analysis of Co-variance

Analysis of co-variance has certain basic assumptions underlying it. According to Ray (1960, *pp*. 129-132), Lindquist (1970, *pp*. 195-196), and Edwards (1971, *pp*. 332−339) the assumptions are that:

1. The distribution of the adjusted scores within groups should be normal.

2. The methods groups should be selected at random from the same population.

3. The groups sholud be homogeneous in variability.

4. The effect to each treatment should be constant and additive on

the responses of the subjects in the group on which the treatment is imposed.

5. There should be a real correlation between the dependent variable and the co-variate or pertinent control variable.

6. There should be homogeneity of regressions from group to group.

The basic assumptions underlying the technique of analysis of co-variance for the data Y_1 and Y_2 in the Table 8.21 are tested as under :

3. *Assumption of normality*. In the light of the findings of Eden and Yates (Johnson, 1961), and Norton (Guilford, 1965, *pp*. 300-301) the assumption of normality may not be considered important for the data of the experiment.

4. *Assumption of randomness*. The requirement of randomness was amply fulfilled for the data of the experiment. First 70 pairs of subjects were chosen on the basis of subject-to-subject matching on the pre-test criterion scores. Finally two groups, with 70 subjects in each group, were framed by randomly assigning one member of each pair to each group.

5. *Assumption of homogeneity*. The assumption of homogeneity of sample variances may be tested by applying Bartlett's test of homogeneity of variance.

1. Computation of ΣY_1K^2 *and* ΣS_1K^2

Using formula (8.12)

$$\Sigma Y_1K^2=\Sigma Y_1{}^2-\frac{(\Sigma Y_1)^2}{N_1}$$

$$=451038-\frac{(5608)^2}{70}$$

$$=1757.09$$

Using formula (8.13)

$$\Sigma S_1K^2=\frac{\Sigma Y_1K^2}{N_1-1}$$

$$=\frac{1757.09}{70-1}=25.47$$

2. Computation of ΣY_2K^2 *and* ΣS_2K^2

Using formula (8.12)

$$\Sigma Y_2K^2=\Sigma Y^2{}_2-\frac{(\Sigma Y_2)^2}{N_2}$$

$$=422299-\frac{(5421)^2}{70}$$

$$=2481.27$$

Using formula (8.13)

$$\Sigma S_2K^2 = \frac{\Sigma Y_2K^2}{N_2-1}$$

$$= \frac{2481.27}{70-1} = 35.96$$

3. *Computation of* $K \log \frac{\Sigma SK^2}{2}$

$$K \log \frac{\Sigma SK^2}{2} = K \log \frac{\Sigma S_1K^2 + \Sigma S_2K^2}{2}$$

$$= 2 \log \frac{25.47+35.96}{2}$$

$$= 2 \log 30.715$$

$$= 2.9744$$

4. *Computation of* $\Sigma \log \Sigma SK^2$

$$\Sigma \log \Sigma SK^2 = \log \Sigma S_1K^2 + \log \Sigma S_2K^2$$

$$= \log 25.47 + \log 35.96$$

$$= 2.9618$$

5. *Computation of difference D*

$$D = K \log \frac{\Sigma SK^2}{2} - \Sigma \log \Sigma SK^2$$

$$= 2.9744 - 2.9618 = 0.0126$$

6. *Computation of chi-square* χ^2

Using formula (1.84)

$$\chi^2 = (2.3026)\ (N-1)\ (D)$$

$$= (2.3026)\ (70-1)\ (0.0126)$$

$$= 2.002$$

7. *Computation of correction C*

Using formula (8.15)

$$C = 1 + \frac{K+1}{3K\left(\frac{N_1+N_2}{2}-1\right)} \quad (K \text{ is the number treatment})$$

$$= 1 + \frac{2+1}{3 \times 2 \times \left(\frac{70+70}{2}-1\right)}$$

$$= 1.007.$$

8. *Computation of corrected chi-square*

Using formula (8.16)

$$\text{corrected } \chi^2 = \frac{\chi^2}{\text{correction } (C)}$$

$$= \frac{2.002}{1.007} = 1.988$$

9. $df = K-1 = 2-1 = 1$

The value of χ^2 needed for significance for 1 df at .05 level is 3.841. Since the obtained value 1.988 of χ^2 is less than the needed value, the variance within the treatments may be treated as homogeneous.

Assumption of additivity. The assumption that the effect of treatment is constant and additive on the responses of the subjects in the group on which the treatment was imposed is tested by the Test of Additivity in the manner described by Ray (1960, pp. 57-58). This test employs a variance ratio calculated in terms of F-value described as under:

1. *Sum of the squares for the experimental group*

$$= \Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1}$$

$$= 451038 - \frac{(5608)^2}{70} = 1757.09$$

Variance for the control group

$$= \frac{1757.09}{N_1 - 1} = \frac{1757.09}{70-1}$$

$$= \frac{1757.09}{69}$$

2. *Sum of the squares for the control group*

$$= \Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2}$$

$$= 422299 - \frac{(5421)^2}{70}$$

$$= 2481.27$$

Variance for the control group

$$= \frac{2481.27}{N_2 - 1} = \frac{2481.27}{70-1} = \frac{2481.27}{69}$$

Variance ratio

$$F = \frac{1757.09}{69} \Big/ \frac{2481.27}{69} = 0.71$$

Since the value of F needed for significance for 69/69 *df* at .05 level of confidence is greater than the obtained value 0.71 of F, the assumption of additivity is found to hold good for the data Y_1 and Y_2 in the Table 8.21.

Assumption of correlation. The assumption that there should be a real correlation between the dependent variable and the pertinent control variable is tested by using the Test of Significance of Regression in the manner described by Ray (1960, p. 129). This test also employs the computation of variance ratio in terms of F-Value. In the present problem, the scores on the intelligence test are the pertinent control variable and those on the criterion test (Post test) as the dependent variable.

The details of the computational steps are described as under:

1. *The predicted sum of squares based on the within sample sums of squares and products*

$$P_w = \frac{\left[\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} + \Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2}\right]^2}{\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2}\right]} \qquad (8.17)$$

$$= \frac{(2563.66 + 1916.10)^2}{(6297.37 + 7386.30)}$$

$$= \frac{(4479.76)^2}{13683.68} = 1466.58$$

2. *The predicted variance*

$$V_{pw} = \frac{Pw}{C-1} \text{ (}C\text{ is the number of groups)}$$

$$= \frac{1466.58}{2-1} = 1466.58$$

3. *Within sample sum of squares of errors of estimate*

$$Ew = \left[\Sigma Y_1^2 - \frac{(\Sigma Y_1)^2}{N_1} + \Sigma Y_2^2 - \frac{(\Sigma Y_2)^2}{N_2}\right] - \frac{\left[\Sigma X_1 Y_1 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} + \Sigma X_2 Y_2 - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2}\right]^2}{\Sigma X_1^2 - \frac{(\Sigma X_2)^2}{N_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N_2}} \qquad (8.19)$$

$$= 1757.09 + 2481.27 - \frac{(2563.66 + 1916.10)^2}{(6297.37 + 7386.30)}$$

$$= 4238.36 - \frac{(4479.76)^2}{13683.67}$$

$$= 2771.78$$

4. *The error variance*

$$V_{ew} = \frac{Ew}{(N-C-1)} \qquad (N = N_1 + N_2)$$

$$=\frac{2771.78}{140-2-1}=20.23 \tag{8.20}$$

5. *The variance ratio*

$$F=\frac{V_{pw}}{V_{ew}} \tag{8.21}$$

$$=\frac{1466.58}{20.23}=72.495$$

Since the obtained value 72.495 of *F* exceeds the value of *F* given in the Table *D* of the Appendix for 1/137 *df* at .05 level of confidence, the assumption of real correlation between the dependent variable and pertinent control variable is justified and that the use of intelligence scores as a pertinent control variable was worthwhile in partially eliminating error from the data.

Assumption of homogeneity of regressions. The assumption that there should be homogeneity of regressions from group-to-group is tested by computing the *F*-ratio for the test of homogeneity of regressions in the following manner as given by Ray (1960, pp. 130-131):

1. *The sums of the squares of errors of prediction*

$$E_i=\left[\Sigma Y_1{}^2-\frac{(\Sigma Y_1)^2}{N_1}\right]-\frac{\left[\Sigma X_1Y_1-\frac{(\Sigma X_1)(\Sigma Y_1)}{N_1}\right]^2}{\left[\Sigma X_1{}^2-\frac{(\Sigma X_1)^2}{N_1}\right]}$$

$$+\left[\Sigma Y_2{}^2-\frac{(\Sigma Y_2)^2}{N_2}\right]-\frac{\left[\Sigma X_2Y_2-\frac{(\Sigma X_2)\,(\Sigma Y_2)}{N_2}\right]^2}{\left[\Sigma X_2{}^2-\frac{(\Sigma X_2)^2}{N_2}\right]} \tag{8.22}$$

$$=1757.09-\frac{(2563.66)^2}{6297.37}+2481.27-\frac{(1916.10)^2}{7386.30}$$

$$=713.42+1984.21=2697.63$$

2. *Within sample sum of squares of errors of estimate* (computed arlier in case of assumption of correlation by using formula (8.19)

$$E_w=2771.78.$$

3. *The measure of homogeneity of regressions*

$$=E_w-E_i \tag{8.23}$$

$$=2771.78-2697.63$$

$$=74.15$$

4. *The variance ratio*

$$F=\frac{E_w-E_i}{C-1}\bigg|\frac{E_i}{N-2C} \tag{8.24}$$

$$=\frac{74.15}{2-1}\Big/\frac{2697.63}{140-2\times 2}$$
$$=3.74$$

Since the value of F needed for significance for 1/136 *df* at .05 level of confidence is greater than the obtained value 3.74 of F, the assumption that two sample regressions are homogeneous is taken to be justified.

Analysis of Co-variance of the Data of the Experimental and Control Groups

After testing the data for the assumptions described above, we may proceed further with the analysis of co-variance to test the significance of the difference between the means of the criterion test scores of the experimental and control groups. The computational steps are given as under:

1. *The first adjustment, based on the sums of squares and products of deviations from the total means*

$$E_t=\left[\Sigma Y^2-\frac{(\Sigma Y)^2}{N}\right]-\frac{\left[\Sigma XY-\frac{(\Sigma X)(\Sigma Y)}{N}\right]^2}{\left[\Sigma X^2-\frac{(\Sigma X)^2}{N}\right]} \tag{8.25}$$

$$=4488.14-\frac{(4352.86)^2}{13748.14}$$
$$=4488.14-1378.18$$
$$=3109.96$$

2. *The second adjustment, based on the within-sample sums of squares and products of deviations from the sample means* by using formula (8.19)

$$E_w=\left[\Sigma Y_1{}^2-\frac{(\Sigma Y_1)^2}{N_1}+\Sigma Y_2{}^2-\frac{(\Sigma Y_2)^2}{N_2}\right]-\frac{\left[\Sigma X_1Y_1-\frac{(\Sigma X_1)(\Sigma Y_1)}{N_1}+\Sigma X_2Y_2-\frac{(\Sigma X_2)(\Sigma Y_2)}{N_2}\right]^2}{\left[\Sigma X_1{}^2-\frac{(\Sigma X_1)^2}{N_1}+\Sigma X_2{}^2-\frac{(\Sigma X_2)^2}{N_2}\right]}$$

$$=1757.09+2481.27-\frac{(2563.66+1916.10)^2}{(6297.37+7386.30)}$$

$$=4238.36-\frac{(4479.76)^2}{13683.67}$$

$$=2771.79$$

3. *The discrepancy sum of squares*

$$D = E_t - E_w \qquad (8.26)$$
$$= 3109.96 - 2771.78 = 338.18$$

4. *The discrepancy variance, a measure of apparent treatment effects*

$$V_d = \frac{D}{C-1} \qquad (8.27)$$
$$= \frac{338.18}{2-1} = 338.18$$

5. *The error variance*

Using formula (8.20)

$$V_{ew} = \frac{E_w}{(N-C-1)}$$
$$= \frac{2771.78}{140-2-1} = 20.23$$

6. *F-value for testing the significance of the difference between means*

$$F = \frac{V_d}{V_{ew}} \qquad (8.28)$$
$$= \frac{338.18}{20.23} = 16.72$$

7. *Degress of freedom (df)*

(*i*) *df* for between samples or treatments $(D) = C-1 = 2-1 = 1$

(*ii*) *df* for within samples or error $(E_w) = -C-N1 = 140-2-1 = 137$

(*iii*) *df* for total means $(E_t) = N-2 = 140-2 = 138$

TABLE 8.22 Summary: Analysis of Co-variance

Components of Variability	*Sum of squares (SS)*	*df*	*Variance (V)*	*F*
Between samples or treatments (D)	338.18	1	338.18	16.72
Within samples or error (E_w)	2771.78	137	20.23	
Total (E_t)	3109.96	138		

8. *The adjusted mean of the criterion test scores of the experimental group.*

$$M_{Y1.X1} = M_{Y1} - b\,(M_{X_1} - GM_X) \qquad (8.29)$$

The adjusted mean of the criterion scores of the aontrol group

$$M_{Y2.X2}=M_{Y2}-b\,(M_{X2}-GM_X) \qquad (8.30)$$

in which

M_{Y1} and M_{Y2} are the means of the scores on criterion test obtained by the subjects of the experimental and control groups respectively.

M_{X1} and M_{X2} are the means of the scores on the intelligence test obtained by the subjects of experimental and control groups respectively.

GM_X is the grand mean of the scores on the intelligence test obtained by the subjects of the two groups.

b is the regression co-efficient

$$=\frac{\left[\Sigma XY-\dfrac{(\Sigma K)\,\Sigma Y)}{\text{N}}\right]}{\Sigma X^2-\dfrac{(\Sigma X)^2}{\text{N}}} \qquad (8.31)$$

$$=\frac{\left[514129-(6471)\,(11029)/140\right]}{\left[312847-\dfrac{(6471)^2}{140}\right]}=0.317$$

Using formulae (8.29) and (8.30):

$$M_{Y1}.X_1=80.11-0.317\,(45.54-46.22)$$
$$=80.33$$
$$M_{Y2}.X_2=77.44-0.317\,(46.90-46.22)$$
$$=77.22$$

TABLE 8.23. Means of Intelligence and Criterion Test Scores of the Experimental and Control Groups

Group	*Number of the Subjects in Group* *N*	M_Y	M_X	*Adjusted Mean* $N_{Y.X}$
Experimental	70	80.11	45.54	80.33
Control	70	77.44	46.90	77.22
		$GM_X=46.22$		

9. *Summary of conclusions.* From the Table *D*, it may be found that the criterion value of *F* for 1 and 137 *df* at .01 level of confidence is approximately 6.83. It is evident from the Table 8.22 that

the obtained value of *F* is 16.72, which is significantly greater than the needed value of significance. Since the obtained *F* value exceeds the value of 6.83, the mean criterion score of the experimental group differs significantly from that of the control group.

Furthermore, it is evident from the Table 8.23 that the adjusted mean of the scores on the criterion test of the experimental group (80.33) is significantly higher than that of the control group (77.22).

Significance of Pearson's Coefficient of Correlation

The mathematical basis for the standard error of a Pearson coefficient of correlation is rather complicated because of the difficulty in its nature of sampling distribution. The sampling distribution of *r* is not normal except when the population *r* is near zero and size of the sample is large ($N=30$ or greater). When *r* is high (0.80 or more) and *N* is small, the sampling distribution of *r* is skewed. This is true also when *r* is low (0.20 or less).

A mathematically more sound method for making the inference regarding Pearson's, especially when its magnitude is very high or very low, is to convert *r* into Fisher's *Z* coefficient and find the standard error (*SE*) of *Z*. The sampling distribution of *Z* coefficient is normal regardless of the size of sample *N* and the size of the population *r*. Furthermore, the *SE* of *Z* depends only upon the size of sample *N*, and is independent of the size of *r*.

The formula for standard error of $Z(\sigma z)$ is

$$SEz-\sigma z=\frac{1}{\sqrt{N-3}}$$

By way of illustration, consider an *r* between the scores on an intelligence test and on an achievement test in arithmetic obtained from a random sample of 84 male students of eighth grade. Using formula (8.32), the standard error of *Z* is

$$\sigma z=\frac{1}{\sqrt{84-3}}=\frac{1}{9}=0.111 \tag{8.32}$$

For an *r* of 0.78, Table *F* of the Appendix gives a corresponding *Z* of 1.05. Since the sampling distribution of *Z* is normal, the confidence interval at .95 level for the population of true $\overline{Z}$ is $Z\pm\sigma z\times 1.96=1.05\ \pm 0{\cdot}111\times 1.96$ or 0.83 and 1.26. The corresponding *r*'s, from the Table *F* of the Appendix, are 0.68 and 0.85, which give a well estimated interval within which we expect the population *r* with

0.95 confidence, that is, the chances are 95 in 100 that population *r* lies between 0.68 and 0.85. If we want a higher degree of confidence we can take the .99 level of confidence, for which the limits are $Z \pm \sigma_{\bar{z}} \times 2.58$ or a confidence interval given by the limits $1.05 \pm .111 \times 2.58$ or 0.76 and 1.34. The corresponding *r*'s from the Table *F*, are 0.64 and 0.87, that is, the chances are 99 in 100 that population *r* lies between .64 and .87.

The Significance of the Difference between Pearson's Coefficients of Correlation

The exact method of determining the standard error of the difference between two *r*'s is first to convert the *r*'s into Fisher's *Z* coefficients and then to determine the significance of the difference between the two *Z*'s.

When we have two correlations between the same two variables, *X* and *Y*, computed from two totally different and unmatched samples, the standard error of a difference between two corresponding independent *Z*'s is computed by the formula

$$SE_{DZ} = \sigma_{Z_1 - Z_2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}} \tag{8.33}$$

in which

N_1 and N_2=sizes of the two samples.

The significance of the difference between two Z's is tested with the help of the formula

$$\bar{Z} = \frac{Z_1 - Z_2}{SE_{DZ}} \tag{8.34}$$

Illustrating the use of the formula (8.34), let us consider the correlations of 0.70 and 0.62 computed between intelligence and academic achievement on the basis of the data obtained from two groups of seventh grade students ($N_1 = 103$) and ($N_2 = 124$) from two different schools. To test significance of the difference between the two *r*'s we shall first convert them into *Z* coefficients. The corresponding *Z*'s, from the Table *F*, are 0.87 and 0.73. Using formula (8.33), the SE_{DZ} is

$$\sigma_{Z_1 - Z_2} = \sqrt{\frac{1}{103 - 3} + \frac{1}{124 - 3}}$$

$$= 0.135$$

$$\bar{Z} = \frac{0.87 - 0.73}{0.135}$$

$$= 1.036$$

The obtained Z value is less than 1.96 and hence is not significant at .05 level. This indicates that the correlation between general intelligence and academic achievement does not really differ in the two groups drawn from two different schools.

STATISTICAL INFERENCE BASED ON NON-PARAMETRIC TESTS

So far we have been dealing with the parametric tests involving the assumptions based upon the nature of the population distribution, and on the way the type of measurement scale is used to quantify the data or observations. More recently there has been the development of another category of tests which do not make numerous or stringent assumptions about the nature of the population distribution. This category of tests is called distribution-free or non-parametric tests.

Non-parametric tests are used when:

1. The nature of the population, from which samples are drawn, is not known to be normal.
2. The variables are expressed in nominal form, that is, classified in categories and represented by frequency counts.
3. The variables are expressed in ordinal form, that is, ranked in order or expressed in numerical scores which have the strength of ranks.
4. The sample sizes are small.

The most frequently non-parametric tests used are discussed as follows:

1. *Unrelated samples*

Chi-square $(\chi)^2$ Test. The chi-square test is used with discrete data in the form of frequencies. It is a test of independence and is used to estimate the likelihood that some factor other than chance accounts for the observed relationship. Since the null hypothesis states that there is no relationship between the variables under study, the chi-square test merely evaluates the probability that the observed relationship results from chance.

The formula for Chi-square (χ^2) is

$$\chi^2=\Sigma\left[\frac{(fo-fe)^2}{fe}\right] \tag{8.35}$$

in which

fo=frequency of occurrence of observed or experimentally determined facts

fe=expected frequency of occurrence.

To evaluate chi-square, we enter Table E of the Appendix with the computed value of chi-square and the appropriate number of degrees of freedom. The number of $df=(r-1)\ (c-1)$, in which r is the number of rows and c the number of columns, in which the data are tabulated.

To illustrate the use of formula (8.35), consider the following data in Table 8.24 of 1000 students who have been categorized into five groups *A*, *B*, *C*, *D* and *E* on the basis of age and their preference for colours blue, green, violet, yellow and red. Across the first row we find that of 200 students falling in group *A*, 75 have given their preference for blue colour, 47 for green, 32 for violet, 30 for yellow and 16 for red. Reading down the column we find that of 298 students given their preference for blue colour, 75 belong to group *A*, 42 to group *B*, 107 to group *C*, 44 to group *D* and 30 to group *E*. The other columns and rows are interpreted in the same way.

TABLE 8.24. The Chi-square Test of Independence in Contingency Table

Colour Group	Blue	Green	Violet	Yellow	Red	Total
A	(59.6) 75	(48.6) 47	(15.6) 32	(40.4) 30	(35.8) 16	200
B	(59.6) 42	(48.6) 41	(15.6) 10	(40.4) 40	(35.8) 67	200
C	(59.6) 107	(48.6) 36	(15.6) 12	(40.4) 22	(35.8) 23	200
D	(59.6) 44	(48.6) 76	(15.6) 16	(40.4) 44	(35.8) 20	200
E	(59.6) 30	(48.6) 43	(15.6) 8	(40.4) 66	(35.8) 53	200
Total	298	243	78	202	179	1000

The hypothesis to be tested is the null hypothesis, namely, that age and colour preference are essentially unrelated or independent. In order to compute χ^2 we must calculate an independence value (expected frequency) for each cell in the contingency table. Independence values are represented by the figures in parentheses within the different cells; they give the number of students whom we should expect to fall in a particular age group, showing their preference for a particular colour in the absence of any real association.

The calculations of independence values (*fe*) and χ^2, are shown as under:

1. *Calculation of independence values* (*fe*)

Row I:

$$\frac{298 \times 200}{1000} = 59.6; \quad \frac{243 \times 200}{1000} = 48.6;$$

$$\frac{78 \times 200}{1000} = 15.6; \quad \frac{202 \times 200}{1000} = 40.4;$$

$$\frac{179 \times 200}{1000} = 35.8$$

Row 2:

$$\frac{298 \times 200}{1000} = 59.6; \quad \frac{243 \times 200}{1000} = 48.6;$$

$$\frac{78 \times 200}{1000} = 15.6; \quad \frac{202 \times 200}{1000} = 40.4.$$

$$\frac{179 \times 200}{1000} = 35.8$$

Row 3:

$$\frac{298 \times 200}{1000} = 59.6; \quad \frac{243 \times 200}{1000} = 48.6;$$

$$\frac{78 \times 200}{1000} = 15.6; \quad \frac{202 \times 200}{1000} = 40.4;$$

$$\frac{179 \times 200}{1000} = 35.8;$$

Row 4:

$$\frac{298 \times 200}{1000} = 59.6; \quad \frac{243 \times 200}{1000} = 48.6;$$

$$\frac{178 \times 200}{1000} = 15.6; \quad \frac{202 \times 200}{1000} = 40.4;$$

$$\frac{179 \times 200}{1000} = 35.8$$

Row 5:

$$\frac{298 \times 200}{1000} = 59.6; \quad \frac{243 \times 200}{1000} = 48.6;$$

$$\frac{78 \times 200}{1000} = 15.6; \quad \frac{202 \times 200}{1000} = 40.4;$$

$$\frac{179 \times 200}{1000} = 35.8$$

2. *Computation of the χ^2 value*

Using the formula (8.35)

$$\chi^2 = \Sigma \left[\frac{(fo-fe)^2}{fe} \right]$$

$$= \frac{(75-59.6)^2}{59.6} + \frac{(47-48.6)^2}{48.6} + \frac{(32-15.6)^2}{15.6}$$

$$+ \frac{(30-40.4)^2}{40.4} + \frac{(16-35.8)^2}{35.8} + \frac{(42-59.6)^2}{59.6} + \frac{(41-48.6)^2}{48.6}$$

$$+ \frac{(10-15.6)^2}{15.6} + \frac{(40-40.4)^2}{40.4} + \frac{(67-35.8)^2}{35.8} + \frac{(107-59.6)^2}{59.6}$$

$$+ \frac{(36-48.6)^2}{48.6} + \frac{(12-15.6)^2}{15.6} + \frac{(22-40.4)^2}{40.4} + \frac{(23-35.8)^2}{35.8}$$

$$+ \frac{(44-59.6)^2}{59.6} + \frac{(76-48.6)^2}{48.6} + \frac{(16-15.6)^2}{15.6} + \frac{(44-40.4)^2}{40.4}$$

$$+ \frac{(20-35.8)^2}{35.8} + \frac{(30-59.6)^2}{59.6} + \frac{(43-48.6)^2}{48.6} + \frac{(8-15.6)^2}{15.6}$$

$$+ \frac{(66-40.4)^2}{40.4} + \frac{(53-35.8)^2}{35.8}$$

$$\chi^2 = 190.05$$

3. $df = (r-1)(c-1)$

$$= (5-1)(5-1)$$

$$= 16$$

The χ^2 critical values for 16 *df* as given in the Table *E* are 26.296 and 32,000 respectively for .05 and .01 levels of significance and the obtained value 190.05 of χ^2 is higher than these values. This indicates that there is a significant relationship between the age and the colour preference and the hypothesis that age and colour preference are essentially independent is rejected at .01 level of significance.

In a 2×2 table which is comprised of 4 cells with $(2-1)(2-1)=1$ df, there is a simple formula that eliminates the need to compute the expected frequencies (independence values) for each cell.

The formula is

$$\chi^2 = \frac{N[\mid AD - BC \mid]^2}{(A+B)(C+D)(A+C)(B+D)} \qquad (8.36)$$

in which

A, B, C and D are the frequencies in the first, second, third and fourth cells respectively and the vertical lines $|AD-BC|$ means that the difference is to be taken as positive.

To illustrate the use of formula (8.36), let us determine whether item 20 of a test differentiates between two groups of boys and girls. The responses to the item are given in the following 2 × 2 Table 8.25.

TABLE 8.25. The Chi-square Test in 2×2 Fold Contingency Table

	Passed item 20	*Failed item 20*	*Total*
Boys	(A) 30	(B) 20	(A+B) 50
Girls	(C) 25	(D) 15	(C+D) 40
Total	(A+C) 55	(B+D) 35	90

Using formula (8.36)

$$\chi^2 = \frac{90\,[\,|(30)\,(15)-(20)\,(25)|\,]^2}{(30+50)\,(25+15)\,(30+25)\,(20+15)}$$

$$= \frac{90\,[\,|450-500|\,]^2}{(50)\,(40)\,(55)\,(35)}$$

$$= 0.058$$

Since the obtained value 0.058 of χ^2 does not exceed the critical value 3.841 of χ^2 necessary to reject the null hypothesis at the .05 level of significance, we may conclude that item 20 of the test does not differentiate between the groups of boys and girls.

When entries in 2 × 2 table are less than 10, Yates' correction for continuity should be applied to formula (8.36). The corrected formula reads:

$$\chi_c^2 = \frac{N\,[\,|AD-BC|-N/2]^2}{(A+B)(C+D)(A+C)(B+D)} \qquad (8.37)$$

The Median Test

The median test is used for testing whether two independent samples differ in central tendencies. It gives information as to whether it is likely that two independent samples have been drawn from populations with the same median. It is particularly useful whenever the measurements for the two samples are expressed in an ordinal scale.

In using the median test, we first calculate the combined median for all scores in both samples. Then both sets of scores at the combined median are dichotomized and the data are cast in a 2 × 2 table shown in Table 8.26.

TABLE 8.26. 2×2 Table for Use of Median Test

	Group I	*Group II*	*Total*
No. of scores above combined median	A	B	A+B
No. of scores below combined median	C	D	C+D
Total	A+C	B+D	

Under the null hypothesis, we would expect about half of each group's scores to be above the combined median and about half to be below, that is, we would expect frequencies A and C to be about equal, and frequencies B and D to be about equal. In order to test this null hypothesis, we calculate, χ^2 using the formula (8.37)

$$\chi^2 = \frac{N\left[\,|AD-BC| - \frac{N}{2}\right]^2}{(A+B)\,(C+D)\,(A+C)\,(B+D)}$$

The following example illustrates the use of median test.

Eighteen male and fourteen female teachers of a co-educational institution were asked to express their attitude towards co-education at the secondary school stage. Both the groups of teachers were administered an attitude scale and common median attitude score was worked out. The number of cases from both groups falling above and below the median score is shown in the Table 8.27.

The null hypothesis that there was no difference between the attitude of the male and female teachers was tested at .05 level of significance using the formula (8.37)

TABLE 8.27. Distribution of Teachers (Male and Female) Below Scores Above the Common Median Attitude Scores

	Below Median	*Above Median*	*Total*
Female Teachers	10	4	14
Male Teachers	6	12	18
Total	16	16	32

$$=\frac{32\left[\,|(10)\,(12)-(6)\,(4)\,|\,-\dfrac{32}{2}\right]^2}{16\times16\times18\times14}$$

$$=3.17$$

Since the obtained value 3.17 of χ^2 for 1 df does not exceed the χ^2 critical value of 3.84 for a two-tailed test at .05 level, the null hypothesis is retained and we may conclude that there is no difference in the attitude of male and female teachers towards co-education at the secondary school stage.

The Mann-Whitney U Test

The Mann-Whitney U test is more powerful than the median test. It is a most useful alternative to the parametric t test when the parametric assumptions cannot be met and when the measurements are expressed in ordinal scale values.

Suppose n_1 is the number of individuals in one of the two independent groups and n_2 the number of the individuals in the other. To apply the U test, we first combine the observations or scores from both groups, and rank these in order of increasing size. In this ranking, we have to consider the algebraic sign, that is, the lowest ranks are assigned to the largest negative numbers, if any. Then the ranks of each sample group are summed individually and represented as ΣR_1 and ΣR_2.

Next we calculate two U's using the formulae:

$$U_1=n_1n_2+\frac{n_1(n_1+1)}{2}-\Sigma R_1 \qquad (8.38)$$

and $$U_2=n_1n_2+\frac{n_2\,(n_2+1)}{2}-\Sigma R_2 \qquad (8.39)$$

in which

n_1=number of individuals in first group
n_2=number of individuals in second group
ΣR_1=sum of ranks in first group
ΣR_2=sum of ranks in second group.

The two U's are also related by the equation:

$$U_1=N_1N_2-U_2$$

Thus only one U need to be calculated, for the other can be easily determined by this equation.

The Z-value of U can be computed by the formula

$$Z=\frac{U-\frac{N_1N_2}{2}}{\sqrt{\frac{(N_1)\,(N_2)\,(N_1+N_2+1)}{12}}} \qquad (8.40)$$

It does not matter which U (the larger or smaller) is used in the computation of Z. The sign of Z will depend on which U is used, but the numerical value will we identical.

The following example illustrates the computation and application of Mann-Whitney U test.

TABLE 8.28. Rating Scores of Student-Teachers Trained Through Micro-teaching and Simulation

Group A	*Rank*	*Group B*	*Rank*
90	39	46	4
78	29	42	1
75	25.5	65	15
72	22	61	12
75	25.5	64	14
83	33.5	82	32
73	23	69	19
80	31	66	16
74	24	56	9
67	17	48	6
63	13	68	18
45	3	44	2
55	8	85	36
84	35	83	33.5
89	38	71	21
77	28	87	37
70	20	76	27
58	10	50	7
47	5	59	11
92	40	79	30
$N_1=20$	$\Sigma R_1=469.50$	$N_3=20$	$\Sigma R_2=350.50$

A researcher wished to evaluate the effectiveness of micro-teaching and simulation in developing certain teaching skills among student-teachers of a teacher training college. He divided all the 40 student-teachers of the college into two groups A and B by randomly assigning 20 to each of the groups. Group A was trained in certainskills of teaching through micro-teaching and the Group B was trained through simulation technique. After a period of two months training, the student-teachers were rated in the teaching skills by supervisors. The rating scores of the student teachers are given in Table 8.28.

All rating scores are ranked from lowest to highest and the Mann Whitney U test is used to test the null hypothesis at the .05 level of significance using the formulae (8.38), (8.39) and (8.40).

$$U_1=(20)\,(20)+\frac{(20)\,(21)}{2}-469.50$$

$$=140.50$$

$$U_2=(20)\,(20)+\frac{(20)\,(21)}{2}-350.50$$

$$=259.50$$

Using the equation $U_1=N_1N_2-U_2$, we check

$$140.50=400-259.50$$
$$140.50=140.50$$

$$Z=\frac{140.50-\dfrac{400}{2}}{\sqrt{\dfrac{(20)\,(20)\,(41)}{12}}}$$

$$Z=-1.61$$

Since the obtained Z value of —1.61 does not exceed the Z critical value of 1.96 for a two-tailed test at .05 level, the null hypothesis is retained and it may be concluded that micro-teaching approach and simulation technique are equally effective in developing certain teaching skills among the student-teachers.

The Kolmogorov-Smirnov Two Sample Test

The Kolmogorov-Smirnov two sample test is a test of whether two independent samples have been drawn from the same population or from populations with the same distribution. If the two samples have been draw from the same population distribution, then the cumulative distributions of both samples may be expected to be fairly close to each other, in as much as they both should show only random deviations from the population distribution. On the other hand, a large enough deviation between the two sample cumulative distributions suggests that the two samples come from different populations.

To use the Kolmogorov-Smirnov two sample test, we make a cumulative frequency distribution for each sample of measurements, using the same interval for both distributions. For each interval then, we subtract one step function from the other. It gives the observed deviations and the test focuses on the maximum of these observed deviations.

If D denotes the maximum of these observed deviations, we have for a one tailed test:

$$D = \text{maximum}\,[Sn_1(X) - Sn_2(X)] \qquad (8.41\ a)$$

and for a two tailed test:

$$D = \text{maximum}\,|\,Sn_1(X) - Sn_2(X)\,| \qquad (8.41\ b)$$

where $Sn_1(X)$ = the observed cumulative step function of one of the samples *i.e.* $Sn_1(X) = K/n_1$, where K is the number of scores equal to or less than X

$Sn_2(X)$ = the observed cumulative step function of the other sample, *i.e.* $Sn_2(X) = K/n_2$

The Kolmogorov-Smirnov Test Applied to Small Samples

When $n_1 = n_2$, and when both n_1 and n_2 are 40 or less, Table G of the Appendix is used in the testing of the null hypothesis. This table gives various values of K_D, which is actually the numerator of D and is defined as the numerator of the largest difference between the two cumulative distributions.

The following study of Lepley (1934) as quoted by Siegel (1956, p. 129) will illustrate the use of the Kolmogorov-Smirnov Test in case of small samples.

Lepley (1934) compared the serial learning of 10 seventh-grade pupils with serial learning of 10 eleventh-grade pupils. His hypothesis was that the primacy effect should be less prominent in the learning of the younger subjects. The primacy effect is the tendency for the material learned early in a series to be remembered more efficiently than the material learned later in the series. He tested this hypothesis by comparing the percentage of errors made by the two groups in the first half of the series of learned material, predicting that the older group of eleventh graders would make relatively fewer errors in repeating the first half of the series than would the younger group of seventh graders.

TABLE 8.29. Percentage of Total Errors in First Half of Series

Seventh-grade Subjects	*Eleventh-grade subjects*
39.1	35.2
41.2	39.2
45.2	40.9
46.2	38.1
48.4	34.4
48.7	29.1
55.0	41.8
40.6	24.3
52.1	32.4
47.2	32.6

The percentage of total errors made in the first half of series by the two groups is presented in the Table 8.29.

To test the null hypothesis that there is no difference in the percentage of errors made in recalling the first half of a learned series between eleventh-grade subjects and seventh-grade subjects, using Kolmogorov-Smirnov Test, the data were cast in two cumulative frequency distributions, shown in Table 8.30., taking $n_1=10$ eleventh graders, and $n_2=10$ seventh graders.

TABLE 8.30. Data in Table 8.29 cast for Kolmogorov-Smirnov Test

Per cent of total errors in first half of series

	24-27	28-31	32-35	36-39	40-43	44-47	48-51	52-55
$S_{10_1}(X)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{5}{10}$	$\frac{7}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$
$S_{10_2}(X)$	$\frac{0}{10}$	$\frac{0}{10}$	$\frac{0}{10}$	$\frac{0}{10}$	$\frac{3}{01}$	$\frac{5}{10}$	$\frac{8}{10}$	$\frac{10}{10}$
$S10_3(X)-S_{10_2}(X)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{5}{10}$	$\frac{7}{10}$	$\frac{7}{10}$	$\frac{5}{10}$	$\frac{2}{10}$	$\frac{0}{10}$

From the Table 8.30, it is evident that the maximum observed deviation between two series is $\frac{7}{10}$. Thus $K_D=7$, the numerator of this maximum deviation. Reference to Table *G* of the Appendix reveals that when $N=10$, a value of $K_D=7$ is significant at .01 level of significance for a one tailed test and we conclude that eleventh-graders make comparatively fewers errors than seventh-graders in recalling the first half of a learned series.

The Kolmogorov-Smirnov Test Applied to Large Samples

When both n_1 and n_2 are larger than 40, we make use of Table *H* of the Appendix instead of Table *G*. While using this table it is not essential that $n_1=n_2$.

For a two-tailed test, we determine the value of *D* for the observed data by using the formula (8.41 *b*). Then we compare this observed value of *D* with the critical value of *D* which is obtained by entering the observed values of n_1 and n_2 in the expression given in Table *H* of the Appendix. If the observed *D* is equal to or greater than that computed from the expression in the Table *H*, we may reject the null hypothesis at the level of significance associated with that expression.

For example, suppose $n_1=40$ and $n_2=70$, and that we wish to make a two-tailed test at .01 level of significance for an observed $D=0.404$. In the row in Table *H* for .01 level, we find the expression for *D* as:

$$1.63\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

Substituting the values of $n_1=40$ and $n_2=70$, we get the value of *D* as:

$$1.63\sqrt{\frac{40+70}{(40)\,(70)}}=0.323$$

Since the observed value 0.404 of *D* is larger than 0.323, we reject the null hypothesis at .01 level.

In case of a one-tailed test, we determine the value of *D* for the observed data by using the formula (8.41 *a*). The probability associated with an observed value of *D* is determined by computing χ^2 using the formula:

$$\chi^2=\frac{4D^2(n_1 n_2)}{(n_1+n_2)}\text{with df}=9 \tag{8.42}$$

The significance of the resulting value of χ^2 with df $=2$ may be determined by reference to Table *E* of the Appendix.

2. *Related samples*

The Sign Test. The sign test is the simplest test of significance in the category of non-parametric tests. It makes use of plus and minus signs rather than quantitative measures as its data. It is particularly useful in the situations in which quantitative measurement is impossible or impracticable, but on the basis of superior or inferior performance it is possible to rank with respect to each other the two members of each pair.

The sign test is applicable either to the case of single sample from which observations are obtained under two experimental conditions and the researcher wishes to establish that two conditions are different or to the case of two equivalent samples in which the subjects are matched with respect to the relevant extraneous variables. The use of this test does not make any assumptions about the form of the distribution of differences. The only assumption underlying this test is that the variable under investigation has a continuous distribution.

Sign Test Applied to Small Samples

If the number of the individuals in the single sample or in each of the equivalent or related samples is less or equal to 25, we make use of the Table *I* of the Appendix to test the significance of the differences between the observations obtained under two different conditions. This table is based on probabilities calculated with the help of binomial distribution.

The null hypothesis tested by the sign test is that

$$P(S_E > S_C) = P(S_E < S_C) = \tfrac{1}{2}$$

in which

S_C and S_E are either the observation scores of the same individual under two different conditions or are the two scores for a matched pair. The null hypothesis, therefore, asserts that the median difference is zero, that is, we expect the number of pairs which have $S_E > S_C$ to equal the number of pairs which have $S_E < S_C$. To test the null hypothesis we find the direction of the differences between every S_{E_i} and S_{C_i} by noting whether the sign of difference is plus or minus and then expect about half of the differences to be positive and half to be negative. The probability associated with the occurrence of a particular number of plus's and minus's can be determined with the help of Table *I* of the Appendix, in which *N* denotes the size of sample and *x* the number of fewer signs.

Let us assume that 10 pairs of matched individuals, were randomly assigned to experimental and control groups. At the end of the experiment the individuals on the basis of their performance, were rated by the judges on 5 point rating scale. Before the experiment it was predicted that the experimental group would be rated better than that of the control group. The results are tabulated in Table 8.31.

TABLE 8.31. Sign Test Applied to Data Consisting of 10 Pairs of Rating Scores Obtained by Experimental and Control Groups

Ratings		*Direction of*	
Control Group	*Experimental Group*	*Difference*	*Sign*
5	2	Sc>Se	+
5	2	Sc>Se	+
4	1	Sc>Se	+
3	4	Sc<Se	–
5	5	Sc=Se	0
3	1	Sc>Se	+
4	3	Sc>Se	+
4	2	Sc>Se	+
5	2	Sc>Se	+
5	2	Sc>Se	+

A rating of 1 represents better performance; a rating of 5 represents poor performance.

Table 8.31 shows one difference in the opposite direction from that predicted, that is, in this case Sc<Se, and, therefore, it received a minus sign. For one pair there was no difference between the two ratings, that is, Sc=Se, and this pair received no sign. The remaining 8 pairs showed differences in the predicted direction.

For the data in Table 8.31, x=the number of fewer signs=1, and N=the number of matched pairs who showed differences 8. Table *I* of the Appendix shows that for N=8, and $x \leqslant 1$ has a one-tailed probability of occurrence under null-hypothesis of $p=0.035$. This value is in the region of rejection for $\alpha=.05$; and therefore, we may conclude that the experimental group as rated by the judges, performed better than that of the control group.

Sign Test Applied to Large Samples

When the number of individuals in the group is larger than 25, the normal approximation to the binomial distribution must be used. The significance of difference is tested by calculating the value of Z given by the formula:

$$Z = \frac{(x \pm .5) - \frac{1}{2}\,N}{\frac{1}{2}\sqrt{N}} \qquad (8.43)$$

where $x+.5$ is used when $x < \frac{1}{2}$ N, and $x-.5$ is used when $x > \frac{1}{2}$N. The significance of the obtained value of Z is determined by reference to Table *A* of the Appendix.

Suppose a researcher is interested in determining whether a film on family planning would change the opinion of the male members of a particular community regarding certain measures to be adopted for having a small family. He draws a random sample of 100 male adults from the community, and conducts a "before and after" study in which each adult serves as his own control. In the start, the researcher asks each individual to express his opinion on whether certain measures may be adopted so as to keep the size of the family small. He then shows the film to the selected 100 adults, after which he repeats the question. The researcher sets the null hypothesis that the film on family planning has no systematic effects on the adults.

The results of this study of the effect of the film upon opinion are shown in Table 8.32.

TABLE 8.32. Opinion of 100 Adults Concerning Certain Measures to be Adopted for having Small Family Before and After Film on Family Planning

		Opinion expressed after a film		
		Favourable	*Unfavourable*	*Total*
Opinion expressed before film	Unfavourable	56	(9)	65
	Favourable	(8)	27	35
	Total	64	36	100

The data in the Table 8.32 show that there were (8+9) =17 adults who were unaffected and (56+27)=83 who were affected. The hypothesis of this study, thus, applies only to these 83. Under the null hypothesis, we would expect about half of those whose opinion changed from before to after to have changed from favourable to unfavourable and about half to have changed from unfavourable to favourable. That is, out of 83, we would expect about 41.5 adults to show each of the two kinds of change. It is evident from Table 8.32 that 56 adults changed their opinion from unfavourable to favourable while 27 changed from favourable to unfavourable. From these data, $x > \frac{1}{2}$ N, that is, $56 > 41.5$. Using the formula (8.43),

$$Z = \frac{(56-0.5)-\frac{1}{2}(83)}{\frac{1}{2}\sqrt{83}}$$

$$= 3.07$$

Using Table *A* of the Appendix, the probability under null-hypothesis of $Z \geqslant 3.07$ is $P=2\ (0.50000-0.4989) = .0022$. The value of P given in the Table *A* is doubled as the region of rejection in the present problem is two-tailed.

Since the value 0.0022 of P is smaller than .01, the null hypothesis is rejected. It may, therefore, be concluded that the film had a significant systematic effect on adults in changing their opinion regarding certain measures which may be adopted for having a small family.

The Wilcoxon Matched-Pairs Signed-Ranks Test

The Wilcoxon matched-pairs signed-ranks test is more powerful than the sign test because it tests not only direction but also magnitude of differences within pairs of matched groups. This test, like the signtest, deals with dependent groups made up of matched pairs of individuals and is not applicable to independent groups. The null hypothesis would assume that the direction and magnitude of pair difference would be about the same.

The application of Wilcoxon test involves the following steps:

1. Let d_i be the difference scores for any matched pair, representing the difference between pair's scores under the two treatments A and B. There would be one d_i for each pair of scores.

2. Delete all such pairs for which $d_i=0$

3. Rank all the d_i's without regard to sign; giving rank 1 to the smallest difference d_i, rank 2 to the next smallest, etc. For example, a d_i of -1 without regard to sign is given a lowest rank than a d_i of either -2 or $+2$. If two or more d_i's are of the same size, assign the same rank to such tied cases. The rank assigned would be average of the ranks which would have been assigned if the d_i's had differed slightly from each other. For example, if three pairs yield d_i's of $-1, -1$, and $+1$, then each pair would be assigned the rank of 2, for $\frac{1+2+3}{3}=2$, and the next d_i in order would get the rank of 4 because ranks 1, 2 and 3 have already been exhausted.

4. Indicate which ranks arose from negative d_i's and which ranks arose from positive d_i's by affixing to each rank the sign of difference.

5. Sum the ranks for the positive differences and sum the ranks for the negative differences. Under the null hypothesis we would expect the two sum to be about equal. In other words, if the sum of the positive ranks equals the sum of the negative ranks, we would conclude that the treatments A and B are different. But if the sum of the positive ranks is very much different from the sum of the negative ranks, we would infer that treatment A differs from treatment B and thus we would reject null-hypothesis.

Wilcoxon Test Applied to Small Samples

When the number of individuals in the single sample or in each of the equivalent or related samples is less or equal to 25, we make use of the Table J of the Appendix to test the significance of the difference between the observations obtained under two different treatments.

This table gives various values of T and their associated levels of significance. The T values indicate the smaller sum of like signed ranks, that is, T is either the sum of the positive ranks of the sum of the negative ranks, whichever sum is smaller. If an observed T is equal to or less than the value given in the Table J under a particular significance level for the observed value of N (N=the number of matched pairs minus the number of pairs whose d=0), then the null hypothesis is rejected at that level of significance.

Table J can be used with both one-tailed and two tailed tests. A researcher must use one tailed test when he can predict in advance the direction of the differences.

The following example illustrates the use and application of the Wilcoxon test.

Ten pairs of identical twins were separated, with one twin placed in a nursery school and the other in an institution for homeless children. After a year they were rated for certain mental tasks and their performance was compared. The data of the experiment are given in the Table 8.33.

The null-hypothesis that there was no difference between the performance of the groups was tested at .05 level of significance. Since the direction of the difference is not predicted, a two-tailed region of rejection is appropriate.

T, the smaller of the sums of the like signed ranks=1+2=3. Since the obtained value 3 of T for N=8 (The total number of pairs 10 minus 2 pairs whose d is zero) is less than the given value 4 of T in

TABLE 8.33. Rating Scores of Identical Twins Assigned to Nursery School and to Institution for Homeless Children

Pairs	*Nursery school (Rating)*	*Institution for homeless children (Rating)*	*d*	*Rank of d*	*Rank with less frequent sign*
a	24	11	13	6	
b	28	14	14	7	
c	26	19	7	3	
d	29	13	16	8	
e	20	20	0	—	
f	22	14	8	4	
g	8	9	—1	−1	1
h	6	6	0	—	
i	23	14	9	5	
j	4	9	—5	—2	2
					T=3

the Table J for a two-tailed test at .05 level, the null hypothesis is rejected and we may conclude that nursery school experience does affect the mental development of children favourably.

Wilcoxon Test Applied to Large Samples

When N is larger than 25, the sum of the ranks is approximately normally distributed. The sum of the ranks of the smaller of the like-signed ranks is designated as T.

The value of Z is determined by the formula:

$$Z = \frac{T - \dfrac{N(N+1)}{4}}{\sqrt{\dfrac{N(N+1)(2N+1)}{24}}} \tag{8.44}$$

in which

N = number of pairs ranked

T = sum of the ranks of the smaller of the like-signed ranks

The associated probability of Z-value under the null hypothesis is determined by referring to Table *A* of the Appendix.

To illustrate the use of formula (8.44), suppose a group of 26 delinquent children were initially rated for their social adjustment by a psychiatrist and then sent to a juvenile jail. After a year they were rated again by a psychiatrist for social adjustment and then their initial and final adjustment rating scores were compared. The rating data are given in the Table 8.34.

TABLE. 8.34. Rating Scores of Delinquent Children

Post-rating score	*Pre-rating score*	*d*	*Rank of d*	*Rank with less frequent sign*
50	47	3	7.5	
54	52	2	5.5	
62	63	—1	—2.5	2.5
39	31	8	17.5	
56	52	4	10	
51	42	9	19	
58	51	7	16	
60	49	11	23	
48	42	6	14	
46	42	4	10	
42	40	2	5.5	
53	56	—3	—7.5	7.5
40	36	4	10	
51	50	1	2.5	
56	42	14	26	
60	48	12	24	
57	47	10	21	
43	48	—5	—12	12
45	37	8	17.5	
52	39	13	25	
61	60	1	2.5	
62	52	10	21	
48	42	6	14	
39	45	—6	—14	14
41	42	—1	—25	2.5
39	49	—10	—21	21

The null hypothesis that there was no difference in initial and final adjustment rating scores of the group was tested at .05 level of significance. Since the direction of the difference is not predicted, a two-tailed region of rejection is appropriate.

T, the smaller of the sums of the like-signed ranks$=2.5+7.5+12.0+14.0+2.5+21=59.5$. Using formula (8.44):

$$Z=\frac{59.5-\dfrac{(26)(26+1)}{4}}{\sqrt{\dfrac{(26)(26+1)(52+1)}{24}}}$$

$$=2.95$$

Since the obtained Z-value of -2.95 exceeds the Z critical value of -1.96 for a two-tailed test at the .05 level, the null hypothesis is rejected. We may conclude that the environment in Juvenile Jail has considerably improved the social adjustment of delinquent children.

The McNemar Test for Significance of Changes

The McNemar test for the significance of changes is particularly applicable to those "before and after" designs in which each individual is used as his own control and in which measurement is expressed in either nominal or ordinal scale. To test the significance of any observed change as a result of a particular treatment, a 2 × 2 fold table of frequencies to represent the first and second sets of responses from the same individuals is set up in the following manner:

TABLE 8.35. 2×2 Fold Table for Use in Testing Significance of Changes

		After	
		—	+
Before	+	A	B
	—	C	D

The signs + and — in the Table 8.35 are used to signify different responses. An individual is tallied in cell A if he changed from + to —. He is tallied in cell D if he changed from — to +. If no change is observed, he is tallied in either cell B (+responses both before and after) or cell C (—responses both before and after).

The significance of the change is tested by calculating chi-square with the help of formula:

$$\chi^2 = \frac{[\,|A-D|-1]^2}{A+D} \qquad (8.45)$$

The significance of any obtained value of χ^2, as computed from formula (8.45), is determined by reference to Table E of the Appendix which gives various critical values of chi-square for df's from 1 to 30. If the obtained value of χ^2 is equal to or greater than that given in Table E for a particular significance level at $df=(2-1)(2-1)=1$, the implication is that a significant change was demonstrated in the before and after responses.

The following example illustrate the application of the formula (8.45).

Suppose a researcher is interested in knowing the effectiveness of some yoga exercises on a group of highly neurotic adults. He sets the hypothesis that use of such excercises, may or may not change the neurotic behaviour of adults. To test this hypothesis, he selects a sample of 25 such adults and rates them initially for their neurotic behaviour. After practising yoga exercises for a period of 30 days, the adults are rated again for their neurotic behaviour. The data thus collected are cast in the form shown in the Table 8.36.

TABLE 8.36. Number of Neurotic Adults Before and After Practising Yoga Exercises

		Number of adults after practising yoga exercises	
		Low	High
Number of adults before practising yoga exercises	High	20	1
	Low	2	2

The Table 8.36 shows that A=20=the number of adults who showed a low tendency in their neurotic behaviour and D=2=the number of adults who showed a high tendency in their neurotic behaviour, after practising yoga. B=1 and C=2 represent those adults who did not show any change in their initial behaviour. Using formula (8.45) for the data in cells A and D:

$$\chi^2=\frac{[\,|20-2|-1]^2}{20+2}$$

$$=13.14.$$

Since the obtained value 13.14 of χ^2 exceeds the χ^2 critical value of 3.84 for a two-tailed test at .05 levels for 1 df., the null hypothesis is rejected; and we may conclude that there is a decrease in the neurotic behaviour of adults after practising yoga.

Significance of Spearman's Coefficient of Correlation rho-coefficient (ρ)

There is no generally accepted formula for estimating the standard error of ρ. We cannot, therefore, determine its confidence limits. We can test the null-hypothesis that the two variables under study are not associated in the population and that the observed value of ρ differs from zero only by chance, in two ways.

1. When the size of sample N is from 4 to 30, the interpretation is best made by the aid of Table L of the Appendix, in which are given ρ co-efficients significant at .05 and .01 levels of confidence. This is a one tailed table, that is, the stated probabilities apply when the

observed value of ρ is in the predicted direction, either positive or negative. For a one-tailed test, if an observed value of ρ equals or exceeds the value of ρ shown in the Table L the observed value is significant at the level indicated. For example, in a study the Spearman's ρ coefficient between length of service and selling efficiency in a group of randomly selected 12 salesmen came out to be 0.76. If our hypothesis is that length of service increases selling efficiency, we have to make use of one-tailed test. Table L shows that a value as large as 0.76 is significant at .01 level of confidence. We, therefore, reject null hypothesis at .01 level, concluding that in the population of salesmen from which the sample was drawn, length of service and selling efficiency are associated.

2. When N is 10 or larger, the significance of on obtained ρ under null hypothesis may by tested by the formula:

$$t=\rho\sqrt{\frac{N-2}{1-\rho^2}} \tag{8.46}$$

For larger N, the value given by formula (8.46) is distributed as Student's t with $df=N-2$. The interpretation of an obtained t by formula (8.46) is made with the use of Table C of the Appendix.

In the previous example, the ρ-coefficient between selling efficiency and length of service came out to be 0.76 for $N=12$. Since N is larger than 10, we may use formula (8.46) for testing the significance of this ρ.

$$t=0.76\sqrt{\frac{12-2}{1-(0.76)^2}}$$

$$=4.86.$$

Table C shows that for $df=N-2=12-2=10$, t as large as 4.86 is significant at .005 for a one-tailed test. This is essentially the same result we obtained previously by usiug Table L. We therefore, reject the null hypothesis at .01 level, concluding that length of service and selling efficiency are significantly related.

Significance of Phi-Coefficient (ϕ-Coefficient)

The significance of ϕ-coefficient is determined through the relationship of ϕ to χ^2 using the formula:

$$\chi^2=N\phi^2 \tag{8.47}$$

This formula enables us to test an obtained ϕ against the null hypothesis. First, we convert ϕ to an equivalent χ^2' and than test the significance of χ^2 by referring to Table E of the Appendix. If χ^2 comes out to be significant for a particular level of confidence, the corresponding value of ϕ is also significant.

Significance of Coefficient of Contingency C

The significance of contingency coefficient C is also determined through its relationship which it bears to χ^2 using the formula:

$$C=\sqrt{\frac{\chi^2}{N+\chi^2}} \tag{8.48}$$

This formula enables us to test an obtained C coefficient against the null-hypothesis by first converting C to χ^2. The interpretation of an obtained chisquare is made with the use of table E of the Appendix. If chisquare is significant at a particular level of confidence, C is also significant.

POST TEST

1. List and describe the assumptions on which the use of parametric test are based.
2. Define the sampling distribution of means.
3. State the characteristics of Central Limit Theorem.
4. Define the standard error of mean.
5. Define the confidence intervals and levels of confidence.
6. Given a sample with mean=25.70 ; σ=4.25 ; and N=100
   (*a*) compute the .95 confidence intervals for the true mean.
   (*b*) compute the .99 confidence intervals for the true mean.
7. Define and illustrate the concept of degrees of freedom.
8. The mean of 12 independent observations of a certain magnitude is 80 and the standard deviation is 14.
   (*a*) compute .95 confidence interval for the true mean.
   (*b*) compute .99 confidence interval for the true mean.
9. The following scores were obtained on a retention test from 6 children:
   6, 7, 5, 8, 6 and 4
   (*a*) Find the .95 confidence interval for the true mean.
   (*b*) Find the .99 confidence interval for the true mean.
10. In a study of abstract reasoning, a sample of 112 seventh grade boys and a sample of 108 seventh grade girls scored as below on a test of abstract reasoning :

Sex	N	Mean	σ
Boys	112	115	14.52
Girls	108	119	9.81

Assuming that our samples are random, is the difference between means significant at .05 level and at the .01 level?

11. An interest inventory is administered in a school to 10 boys whose conduct records are exemplary, and to 8 boys whose records are poor. Data are given below :

Group 1 (Boys) : 115, 117, 100, 110, 116, 102, 117, 107, 109, 113.
Group 2 (Girls) : 105, 107, 90, 100, 108, 114, 114, 101, 104.

Is the difference between group means significant at the .05 level and at the .01 level?

12. Two groups, 1 and 2, of eighth grade children, 82 in each group, are paired child for child for age and score on Form A of a test of intelligence. Five weeks later, both groups are given form B of the same test. Before the second test, Group I, the experimental group, is praised for its performance on the first test and urged to try to better its score. Group 2, the control group, is given the second test without comment. Will the incentive (praise) cause the final score of Group I and Group 2 to differ significantly ? The relevant data is given as under :

Number of children	*Group 1 (Experimental)*	*Group 2 (Control)*
Mean score on form A (initial test)	70.42	70.51
SD on form A (initial test)	25.61	25.57
Mean scores on form B (final test)	77.26	68.30
SD on form B (final test)	26.23	20.36
Correlation between final scores (experimental and control groups)	0.70	

13. Ten subjects are given 4 successive trials upon a digit symbol test, of which only the scores for trials 1 and 4 are shown below. Test the hypothesis that practice increases the test scores from initial to final trials significantly.

Trial 1 : 60, 52, 61, 36, 45, 52. 70, 51, 80, 65, 72, 48.
Trial 4 : 72, 50, 71, 45, 40, 62, 78, 61. 94, 73, 82, 60.

14. Two groups of elementary school children are matched for mean and standard deviation upon a group intelligence test. There are 40 subjects in Group 1 and 60 subjects in Group 2. The records of these to groups upon a battery of 'learning' test are as follows:

	Group 1	*Group 2*
Mean	38.42	43.51
Standard deviation	8.60	13.35
Number	40	60

The correlation of the group intelligence test and the learning battery in the entire group from which the groups 1 and 2 are drawn is 0.55. Is the difference between the two groups significant at the .05 level and at the .01 level?

15. Describe the uses of analysis of variance.
16. List and state the basic assumptions underlying the technique of analysis of variance.
17. Twenty pupils are randomly assigned to 4 groups of 5 each, which are taught by different methods. At the end of the experiment same test of achievement is administered to all the pupils.

    The scores on this test are given in the following table.

Methods or Groups			
A	*B*	*C*	*D*
14	19	12	17
15	20	16	17
11	19	16	14
10	16	15	12
12	16	12	17

    (*a*) Test the data for the assumptions of analysis of variance.
    (*b*) Apply the analysis of variance to test whether the groups differ from one another in the mean achievement scores.

18. In a 2×2×2 factorial experiment, ten subjects are assigned at random to each treatment combinations of A, B, and C. The data are given in the following table:

A_1				A_2			
B_1		B_2		B_1		B_2	
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
7	6	10	3	10	1	1	9
7	8	1	2	2	4	10	6
4	1	3	2	8	3	10	5
1	7	9	5	10	4	6	6
5	7	9	2	3	6	7	3
9	7	4	4	1	9	3	9
6	8	1	7	4	3	5	8
8	7	3	1	3	4	8	3
9	9	10	8	4	4	3	7
8	3	3	9	1	10	5	2

    (*a*) Test the data for the assumptions of analysis of variance.
    (*b*) Apply the analysis of variance to test the significance of (*i*) main effects and (*ii*) interaction effects.

19. Describe the uses of analysis of co-variance.
20. List and state the basic assumptions underlying the technique of analysis of co-variance.

21. Suppose 15 children have been given one trial X of a test. Five are then assigned at random to each of the three groups A, B and C. After two weeks, group A was praised lavishly, group B scolded severely, and the test was repeated (Y). At the same time, a second trial (Y) is also given to group C, the control group, without comment. The data are presented in the following table:

Group A	*(Praised)*	*Group B*	*(Scolded)*	*Group C*	*(Control)*
X_1	Y_1	X_2	Y_2	X_3	Y_3
20	40	35	38	15	20
15	30	20	22	20	25
25	35	25	30	30	30
10	25	25	20	15	20
15	30	20	20	20	20

(*a*) Test the data for the assumptions of analysis of co-variance.

(*b*) Apply the analysis of co-variance to test the significance of the difference between the mean scores on the test Y among the three groups.

22. The Pearson's coefficient of correlation (r) of 0.80 is obtained from a random sample of 50 cases.

(*a*) compute .95 confidence interval for the true r.

(*b*) comptue .99 confidence interval for the true r.

23. The Pearson's correlation coefficient between height and weight of 100 ten-year-old boys is 0.65 and the correlation between height and weight in a sample of 150 ten-year-old girls is 0.60. Is the difference between the correlations significant at 0.05 level and at .01 level?

24. List and describe the assumptions on which the use of non-parametric tests are based.

25. Describe the use use of chi-square test.

26. The table below shows the number of normals and abnormals who chose each of the three possible answers to an item on a neurosis questionnaire.

	Yes	No	?	Total
Normals	15	80	20	115
Abnormals	27	80	8	115
	42	160	28	230

Apply chi-square to test whether this item differentiates between the two groups?

27. Describe the use of median test.

28. In answering a questionnaire the following scores were achieved by 10 men and 20 women.

Men : 12, 21, 28, 37, 38, 38, 39, 40, 42, 51
Women : 12, 13, 15, 15, 21, 23, 24, 24, 25, 27, 30, 31, 32, 33, 34, 36, 38, 43, 44.

Using a median test, test the null hypothesis that the scores of men and women did not differ significantly at .05 level.

29. Describe the use of Mann-Whitney U-test.
30. A group of 50 school students was randomly assigned to experimental and control groups to determine the effectiveness of a counselling programme upon academic averages. Using the Mann-Whitney U test, test the null hypothesis that there was no difference between the academic performance of the experimental and control groups at the .05 level.

Experimental	*Control*
2.12	2.04
3.06	2.68
2.06	3.07
2.14	2.24
3.24	2.76
3.60	2.47
2.55	3.41
1.68	4.01
1.96	2.99
2.04	2.57
2.92	2.79
3.12	2.11
3.71	2.74
2.68	3.36
2.58	2.83
2.52	3.07
3.79	2.69
2.42	1.92
3.22	1.72
1.73	1.59
3.06	1.41
2.08	2.11
2.88	3.70
3.06	2.13
1.90	2.86

31. Describe the use of Kolmogorov-Smirnov two sample test.
32. The following data represents the percentage of errors made by the two groups of students in repeating the first half of the series of a learned material.

Group 1	*Group 2*
39.1	35.2
41.2	39.2
45.2	40.9
46.2	38.1
48.4	34.4
48.7	29.1
55.0	41.8
40.6	24.3
52.1	32.4
47.2	32.6

Using Kolmogorov-Smirnov two sample test, test the null hypothesis that the percentage of errors made by the two groups did not differ significantly.

33. Describe the use of sign test.
34. Twenty-six members of the Weight Watchers Club were weighed three months after the conclusion of the enrolment period. Using sign test, test the hypothesis that the programme was not effective in enabling the former members to sustain their weight loss achieved in the programme.

Initial weight	*Two months after*
157	146
173	168
168	158
198	187
158	161
138	142
123	118
202	198
226	212
268	248
163	162
147	151
120	118
128	142
148	146
146	143
127	118
178	167
173	168
166	163
192	188
210	204
233	218
158	163
198	178
243	248

35. Describe the use of Wilcoxon Matched-Pairs Signed-Ranks Test.
36. At random one twin form each of the eight pairs of identical twins was admitted to a nursery school and the other of the pair was placed in his own home. At the end of the term, the 16 children were given the test of social perceptiveness. The data are presented in the following table:

Pair	*Social perceptiveness score of twin in nursery school*	*Social perceptiveness score of twin at home*
a	72	53
b	59	32
c	63	64
d	33	27
e	48	41
f	46	33
g	66	70
h	55	72

Using Wilcoxon Matched-Pairs Signed-Ranks Test, test the hypothesis that the social perceptiveness of home and nursery school children does not differ at .05 level of significance.

37. Two groups of high school students were matched on the basis of intelligence quotients. One group was taught a unit on the advantages of yoga by lecture-discussion method; the other by a prediscussion and film strip presentation. At the conclusion of the unit the groups were tested. Test the null hypothesis that there was no difference in understanding between the groups at the .05 level of significance. Use the Wilcoxon Matched-Pairs Signed-Ranks Test. The data are given as under:

Subject	*Lecture*	*Filmstrip*
1	50	60
2	78	88
3	68	69
4	72	70
5	50	65
6	44	60
7	70	77
8	44	58
9	40	68
10	70	79
11	50	74
12	80	86
13	60	78
14	76	64
15	50	68
16	65	55
17	78	82
18	50	70
19	82	60
20	86	78

38. Describe the use of McNemar test for significance of changes.
39. A nursery school teacher observes 25 children on each child's first day at nursery school, and he categorizes their first initiation of social contact according to whether it was directed to an adult or to another child. He then observes each of the 25 children after each has attended nursery school for a month, making the same categorization. The data he obtains is presented in the following table:

		Object of initiation on thirtieth day	
		Child	Adult
Object of initiation on first day	Adult	14	3
	Child	5	3

Using McNemar Tast, test the hypothesis that with increasing familiarity and experience, children have gained in the initiation of their social contacts with other children rather than with adults.

40. The Spearman's coefficient of correlation between length of service and selling efficiency in a group of 12 salesmen has come out to be 0.80. Test the significance of the correlation at .05 and .01 levels.
41. Two items X and Y are part of a test of 100 items. Item X was passed by 100 and failed by 125 students in a group of 225. Item Y was passed by 135 and failed by 90 in the same sample. The phi-coefficient of correlation between X and Y came out to be 0.36. Test the significance of phi-coefficient by way of χ^2.
42. The contingency coefficient of correlation between the marriage adjustment score of 500 husbands and their education in a study came out to be 0.24. Test the significance of this correlation by way χ^2.

9. The Research Report

This chapter deals with the procedure followed in the preparation of the research report. It has four sections. The first section outlines and describes in detail the general format of the research report under three sub-sections : the preliminary section, the body of the report or text, and the reference section. The second section describes briefly the style and format of writing the research report logically and objectively. The third section describes in detail the procedure which is followed in typing different sections of the research report. The fourth and the final section provides certain hints which are useful in detecting and correcting the typing errors while making proof reading of the final research report.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following Objectives:

1. Describe in detail the general format of the research report.
2. Describe the style and format of writing the research report logically and objectively.
3. Write down the standard abbreviations of some words which are generally used in the footnotes, the tables and the bibiliography of the research report.
4. Describe in detail the procedure which is followed in typing different sections of the research report.
5. Describe the procedure which is adopted in proof reading of the final draft of the research report.

The research is obligated to give a completely detailed account of all his experience in the thinking process involved in the method of the research so as to carry the reader with him from identification and definition of the problem, formulation of hypotheses, collection of data as evidence, analysis and interpretation of data, to testing of hypotheses and drawing out of conclusions. Such a complete and detailed account of research experience is called a research report. A comprehensive coverage of research experience and an appropriate communication of the results may be possible only when the researcher possesses requisite knowledge of the procedures involved in writing a research report. A knowledge of the rules and principles involved in writing a research report will help the researcher to write better theses, dissertations and journal articles.

GENERAL FORMAT OF THE RESEARCH REPORT

For the purposes of presentation of a research report, several style manuals are available which provide guidance to the researcher as to the specific rules on style and format to be followed in reporting findings of his research. Some universities, research organizations or journal boards have established their own format to which their theses, dissertations or research papers must conform, but all formats are somewhat similar to the following outline which comprises of three main sections: (1) the preliminary section; (2) the body of the report to text and (3) the reference section. Each main section consists of several sub-sections.

Preliminary Section

1. Title page
2. Preface, including acknowledgements (if necessary)
3. Table of contents
4. List of tables (if any)
5. List of figures, maps or illustrations (if any).

Body of the Report or Text

(*a*) **Introduction**

1. Statement of the problem
2. Analysis of previous research
3. Relation of present problem to theoretical position of the previous research
4. Significance of the problem
5. Delimitations of the study
6. Assumptions underlying hypotheses
7. Statement of hypotheses
8. Definition of important terms.

(*b*) **Design of the study**

1. Procedures employed
2. Sources of data
3. Data gathering instruments
4. Sampling and methods of gathering data.

(*c*) **Analysis and interpretation of the data**

1. Text
2. Tables (if any, are usually included into the text)
3. Figures (if any, are usually included into the text).

(*d*) Summary and conclusions

1. Brief restatement of problem and procedures
2. Principal findings and conclusions with their practical implications (if any)
3. Suggestions for further research.

Reference Section

1. Bibliography
2. Appendix
3. Index (if any).

Preliminary Section

1. *Title page.* Several pages of preliminary material are presented prior to the body of a research report. Generally it contains the following information:

(*i*) Title of the study.
(*ii*) Name of the institution to which the report is being submitted.
(*iii*) Degree for which the report is submitted.
(*iv*) Name of the candidate (if desired, previous academic degrees may be listed after name)
(*v*) Date of submission of the report.

2. *Preface.* The preface usually include a brief statement of the purpose and scope of the report. A preface should also include thanks for those who gave the researcher substantial guidance or assistance in the conduct of the study. If the researcher has little to say about the contribution of his research work, the preface can easily be omitted. In such a situation, the page should be labelled 'Acknowledgements' rather than 'Preface'. Acknowledgements should be brief and simple. A long list of effusive Acknowledgements full of flattery is not in good taste. The brief statement of acknowledgement should sincerely recognize the persons and institutions to whom the researcher is indebted for providing academic guidance, administrative facilities and financial assistance.

3. *Table of contents.* The table of contents includes the major divisions of the report; the introduction, the chapters with their sub-sections, the bibliography and the appendix. Page numbers for each of these divisions and sub-divisions are given. The titles of chapters and sub-divisions within chapters should correspond exactly with those included in the body of the report with correct page citation. The preface or acknowledgements, list of tables and list of figures are also entered in the table of contents. The table of contents provides

reader a birds' eye view of the report and enables him to locate quickly each section of it.

4. *List of tables and figures.* It tables and figures are presented in the research report, a separate page is included for each list. These lists include full titles of tables and figures with their corresponding numbers and page citation.

Body of the Report

The main body of the research report usually contains four logical divisions: (1) an introduction; (2) design of the study, (3) analysis and interpretation of the data; and (4) summary and conclusions. In a thesis or a dissertation these divisions comprise chapters, while in a research paper they may consist of sections with suitable headings.

1. *Introduction.* The introduction of a research report should be lucid, complete and concise. It should introduce the research problem in a proper context, and arouse and stimulate the reader's interest. If introductions are dull, confused, and lacking in precision, direction and specificity, there is little incentive for the reader to continue reading the research report.

In the introduction the researcher defines, analyses and states the nature of the problem. He also reviews the related studies so as to lay a foundation for research. The review of the results of previous researches brings out areas of agreement and disagreement, and shows how the present study arose from contradictions or inadequacies of earlier investigations. The introduction also includes the significance of the problem and the need for conducting the investigation. After reviewing the background of the problem, its scope and delimitations, the researcher presents his hypotheses and the assumptions on which the hypotheses are based. He then defines the terms used in the study that have a special meaning or significance for the investigation.

2. *Design of the study.* This section explains the design of the study in detail. It includes a detailed description of the manner in which decisions have been made about the type of data needed for the study, the tools and devices used for their collection, and the method by which they have been collected. A researcher may present a definition of the population; the size of the sample and the rationale for the size; the method of sampling; the number of individuals who declined to participate. weeded out, or did not participate in different phases of study and why, where, when and what types of data were collected; the tools and devices used for collecting data along with their reliability and validity; the design and method of conducting the

expriment giving full details about the assumptions, classification and manipulation of variables, and nature of treatments; directions given to the subjects; the characteristics of interviewers or observers and the type of training provided to them; the types of data analyses made; the statistical methods employed and reasons for selecting such methods; and how the data will be organized and presented for analysis and interpretation.

3. *Analysis and interpretation of the data.* This section is the heart of the research report. The data analysis and interpretation may either be presented in separate chapters or may be integrated and presented in one chapter. The data is presented in tables and figures accompanied by textual discussion. The tables and figures should be constructed and listed in such a way that they clarify significant relationships and become self-explanatary. Tables which are complex and lengthy should be placed in appendix, otherwise the continuity of textual discussion gets broken. In the textual discussion of the data, the report should not repeat all the detailed information that is provided in the tables and figures. He should only point out important facts and relationships to give meaning to the data and make certain generalizations about the data.

The formulae and statistical procedures which were used in the analysis of the data should be clearly specified and explained in detail. Statistical information may be presented in the form of sums, ratios, proportions or percentages; frequency distributions; means and variances or standard deviations; coefficients of correlation, and prediction. The reasons for selecting a particular test of significance, the assumptions underlying its use and the confidence levels chosen in arriving at the results must be presented carefully. All the unexpected developments in the form of unanticipated relationships or unforeseen trends should be reported fully. Any weakness in the research design, tools, techniques, or population that have come to light during the conduct of study should be discussed frankly along with the manner in which factors may have affected the findings of the study.

4. *Summary and conclusions.* This section includes a brief re-statement of the problem, a description of the procedures used, and discussion of findings and conclusions of the study. The conclusions are presented concisely and related directly to the hypotheses that were tasted. They announce whether the findings of the study accept or reject the hypotheses. Conclusions are answers to the questions raised and suggest modification in the existing theory. In addition, the

researcher may list unanswered questions that have occurred in the process of study and which require further research beyond the scope of the problem investigated. If no further research would appear to be advantageous in the area investigated and a new approach to the problem is needed, the researcher should make suggestions. In short, the discussion and presentation of the conclusions should leave the reader with the impression of completeness and of positive gain.

The summary and conclusion section is the most widely used part of the research report because it reviews all the information that has been presented in its previous sections. Most readers scan this section of the report first to get an overview of the study and to determine its utility to them. If the study is of some utility to them, they go through the remaining chapters also.

In addition to summary, some institutions and research journals require a researcher to submit an abstract of study. This abstract serves as a synopsis that gives relevant information to a reader to judge whether he wishes to read the complete research report.

Reference Section

The reference section includes bibliography and appendix. The bibliography, which will be discussed in detail later in this chapter, follows the main body of the report. In a research article, the heading 'references' is used in place of bibliography. The bibliography is a record of those sources and materials that have been used for the study If the number of references is large, the researcher may divide the bibliography into various sections, one for books, one for periodicals and journals, and possibly one for reports and special documents.

An appendix follows the bibliography. All the relevant supporting unwiedly materials, that are important but not essential to the understanding of the report, are presented in the appendix. These materials include questionnaries, copies of covering letters used, evaluation sheets, checklists, courses of study, long quotations, documents, tests, interview forms, and raw data.

If a study is of major importance and is to be published in book or monograph form, the researcher also prepares an index in alphabetical order which follows the appendix.

STYLE AND FORMAT OF WRITING

Language

An outstanding scientific investigation is of little value if its process and findings are not communicated effectively to others. Compbell

(1969), Dugdale (1967) and Turabian (1967) have provided several style manuals. After adopting any of the style manuals, the researcher must adhere to it throughout the report. A clear statement of the hypotheses, and logical and objective presentation of the procedure rather than emotionalized discriptions are required in a scientific report. A research report is usually read by well informed readers who always search for flaws in the process of reasoning. It is possible that the readers may check the results by repeating experiment or computations. To meet this eventuality, a research report must be presented so carefully that it can stand the test of critical scholarship of other readers.

The presentation of the research report should be creative, logical and concise making use of simple common words and sentence structure whenever possible. Its language should be formal and straight forward avoiding slang, proverbial or discourteous phrases. The personal pronouns I, we, you, my, our, and us should not be used. For example, instead of saying, "I administered achievement test in general science to both the groups of students", the passive voice construction. "Achievement test in general science was administered on both the groups of students", would be preferable. The personal pronouns may be avoided by the use of such expressions as the "the researcher" or "the investigator.

The use of abbreviations, except some universally acceptable ones such as IQ, M.A., etc. should be avoided in the main text of the research report. In the footnotes, the tables and the bibliography, some standard abbreviations are used to conserve space. A researcher should be familiar and should master the following standard abbreviations:

bk., bks.	book, books
chap., chaps.	chapter, chapters
p., pp.	page, pages
anon.	anonymous
cf.	compare
col., cols.	column, columns
e.g.	for example
edn., edns.	edition, editions
ed., eds.	editor, editors
et al.	and others
Ibid.	same reference
Op. cit.	previously cited
Loc. cit.	the place cited
fig., figs.	figure, figures

Idem.	same person
illus.	illustrated
l., ll.	line, lines
f., ff.	and the following page (s) (pp. 5f, Page 5 and the following page. pp. 5 ff., Page 5 and the following pages).
mimeo	mimeographed
MS, MSS	manuscript, manuscripts
n , nn.	note, notes
N.B.	please note
n.d.	no date
n.n.	no name
n p.	no place
N.S.	new series, new style
no., nos	number, numbers
pt., pts.	part, parts
sec., secs.	section, sections
trans.	translator, or translated by
viz.	namely
vol. ols.	volume, volumes
vs. vss.	verse, verses
rev.	revised, or revision
passim.	scattered
Sic.	thus
Vide.	see

Numbers of less than three figures, round numbers and numbers that begin sentences are spelled out, except in statistical discussions in which they are used frequently. Fractions are also spelled out unless they are part of longer numbers. Figures are used for all the small and large numbers which appear in a series. In numbers with more than three digits, commas are usually used to point off thousands or millions. e.g. 2, 503; 21, 114, 301. Use either letters or numbers in parentheses to enumerate items in a paragraph. Both forms of letters and numbers in parentheses are not used in the same report. All numbers beginning sentences are spelled out. Figures are used for decimals and percentages, but the word 'per cent' is spelled out. *e.g.* 12 per cent.

To ensure continuity of the text, neither the standard statistical formulae nor the computations are included in the research report. If a rather unusual formula is used in the analysis of data it is appropriate to include it.

The ordinary rules of correct usage should prevail throughout the research report.

Best (1977, p. 318) points out that he has:

> "noted frequent errors of spelling, agreement between subject and predicate, non-parallel construction, and inconsistent tense sequence in the research report."

The researcher should make use of a handbook of style (Compbell, 1969; Dugdale, 1967; Turabain, 1967), a good dictionary, a spelling guide and Roget's Thesaurus while writing the research report. He should use past tense when referring to what the researcher or other researchers have done. The present tense should be used when the researcher is referring the reader to the tables and charts that are presently before him and when he is presenting general truths and well-established principles.

TYPING OF THE RESEARCH REPORT

A typist with great proficiency should be engaged for typing the research report. It is the responsibility of the researcher to present manuscript material to the professional typist in proper form. The manuscript should adhere strictly to the institutional requirements to which the report in the form of a thesis, dissertation or research article is to be presented. A well and accurately typed report makes its presentation interesting and meaningful to the reader. The following are some rules which should be followed while typing the research reports.

General Rules for Typing Research Report

1. A white bond paper of 8½ by 11 inches size of 13 to 16 pound weight should be used for the original and first carbon copy of the thesis or dissertation. A lighter weight paper may be used for other carbon copies. Any bond paper is acceptable for typing the research articles.
2. Material should be typed on one side of the paper.
3. A typewriter with large (pica) type should be preferred. Same style and size of type should be used throughout the report.
4. Use a non-greasy, fresh and black carbon paper, a medium-inked type writer ribbon, and a clean type to get a clear and dense copy.
5. The right margin should be one inch, the left margin 1½ inches, the top margin 1¼ inches, and the bottom margin 1½ inches.
6. All textural material should be double spaced.

7. Throughout the manuscript an indention of seven spaces should be used at the beginning of paragraphs and quotations.

8. Dividing of words at the end of the line should be avoided as far as possible.

9. The body of the text is normally double spaced. Different spacing is required for quotations, footnotes, tables and figures, appendices and index (if any). These requirements are outlined with the discussion of each of these separately.

Rules for Typing Different Sections of Research Report

The pattern and rules for typing preliminaries, main tent, tables, footnotes, quotations, bibliography, appendix and index ((if any) vary slightly from each other. A detailed discussion of each of these will help the researcher to provide the typist with the necessary directions for producing a satisfactory type-script.

1. *Title page*. Most universities and research organizations prescribe their own form of title page and these should be complied within all matters of content and spacing. The title should by typed in capital letters beginning six spaces from the top of the page. If the title is too long to be centred on one line, an inverted pyramid style should be followed, without splitting words or phrases. It is not a good practice to underline titles and include them within inverted commas. Below the title, other items may be centred or balanced against the left and right margins of the page. Other items are not typed in the capital letters, but usually only the initials of the principal words are capitalized.

2. *Preface or acknowledgements page*. The heading 'Preface' or 'Acknowledgements' in capitals should be centred Three spaces below-it should begin the first line of its contents. The researcher's initial are placed three spaces below the last line of the contents of acknowledgement and to the right of the centre of the page.

3. *Table of contents*. The heading 'Table of Contents' in capitals should be centred at the top of the page. About three spaces below this headings 'page' appears at the right margins. Below this are preface or acknowledgements, list of tables and list of figures. Then the heading "chapter" appears at the left margin. The initial letters of column headings "chapter" and "page" are in the capital form. The 'preface' or 'acknowledgements', 'list of tables' and 'list of figures' along with the chapter headings should be typed in capitals without terminal punctuation. The chapter headings are numbered consequently in roman capitals. The initial letter of the first word of

sub-headings and of all nouns, pronouns, verbs, adjectives and adverbs should be capitalized. To insert page numbers for each division and sub-division, it is advisable to type the table of contents only after the entire final draft of the research report is typed.

The setting out of a sample table of contents is shown on page 354.

CONTENTS

Chapter	*Page*
PREFACE (ACKNOWLEDGEMENTS)	*ii*
LIST OF TABLES	*iii*
LIST OF FIGURES	*iv*
I. *Introduction*	
Statement of the Problem	1
Analysis of the Previous Research	1
Relation of Present Problem to the Theoretical Position of the Previous Research	12
Significance of the Problem	19
Assumption underlying Hypotheses	22
Statement of Hypotheses	24
Objectives and Delimitations of the Study	26
Definition of Important Terms	29
II. *Chapter Title*	
(Sub divisions as needed)	33
Bibliography	
Appendices	
Appendix A	
Appendix B	
Index	

An alternative method of listing the table of contents is based on the use of numbering system. For example, Chapter IV of the sample table of contents would be set as follows:

IV. *Analysis and Interpretation of Data*	
4.1. Need Achievement	64
4.11 Test of Homogeneity	65
4.12 Results of Analysis of Variance	67
4.13 Main Effects	75
4.14 Interactions	78
4.2. Need Defference	
4.21 Test of Homogeneity	85
4.22 Results of Analysis of Variance	90
4.23 Main Effects	93
4.24 Interactions	105

4. *List of tables and figures.* Following the table of contents are separate pages for the list of tables and the list of figures with headings, LIST OF TABLES and LIST OF FIGURES respectively Two spaces below these headings *Table* and *Page* (in case of list of tables) and *Figure* and *Page* (in case of list of figures) appear at the left and right margins respectively. The information for each item in these lists includes number of the table of figure in arabic numerals, the exact title of the table or figure, and the page number on which it appears in the main body of the research report. The initial letters of key words in titles are capitalised and no terminal punctuation is used. For illustration, the samples of the list of tables and figures are presented below.

LIST OF TABLES

Tables		*Pages*
1.	Means and Standard Deviations of Raw Scores on General Mental Ability Test of Urban High School Boys	35
2.	Means and Standard Deviations of Raw Scores on General Mental Ability Test of Urban High School Girls	42
3.	Pairwise Comparision of Means of Boys and Girls at different Levels of General Mental Ability	95
4.	Summary of the Results of Analisis of Variance	112

LIST OF FIGURES

Figures		*Pages*
1.	Sketch illustrating the Treatments at Three Levels of General Mental Ability and Socio-economic Status	55
2.	Illustration of Phenomenon of Interaction between Socio-economic Status and Sex in Need Achievement	63
3.	First Orthogonal Rotation of Axes	97
4.	Second Orthogonal Rotation of Axes	116

5. *Chapter divisions and sub-divisions.* For clarity and ease of reading, the chapters of these and dissertations are usually divided into chapter divisions and sub-divisions. This practice is not followed in research articles.

The method of labelling chapter divisions and sub-divisions depends on the number of such divisions to be made. The kind of headings used include centre headings, side headings, and paragraph headings. In addition, each chapter has a chapter number and chapter heading.

Usually, centred headings are used for chapter divisions, and side and paragraph headings for sub-divisions.

Many combinations of headings are used. The commonest types of combinations used are three, four, and five levels of sub-division within a chapter. The following examples illustrate these levels of sub-division within a chapter.

There levels of sub-division

Centre Heading

Side Heading

Paragraph heading

Four levels of sub-division

Centre Heading

Side Heading

Paragraph heading
Paragraph heading

Five levels of sub-division

Centre Heading

SIDE HEADING

Side Heading

Paragraph heading
Paragraph heading

Once a method of headings and sub-headings has been adopted, this method should be consistently used from chapter to chapter throughout the report. To maintain a consistent style, the following rules may be followed in typing and spacing of chapter headings, centre headings, side and paragraph headings.

Chapter headings. Each chapter begins on a new page even if the last page of the preceding chapter has only two or three lines. The chapter number is typed in capitals and roman numbers with no punctuation. It is centred and placed two inches lower from the top of the page. The chapter title is also centred and typed in capitals with no punctuation. It is centred horizontally one double-space below the chapter number. The next heading or textual material follow three spaces below the title.

Example:

↑ Two
| inches from top of page
CHAPTER IV double space
TITLE OF CHAPTER IV Triple space
Next heading or textual material

Centre headings. Centre headings are used for major chapter divisions. It is numbered and centred horizontally on the page and typed in lower case except for the initial letter of key words. Three single spaces usually separate the centre headings from text or from other headings above and below.

Side headings. Side heading is not numbered. Three single spaces separate it from the text above and one double space from the text below. Side heading is also typed in lower case except for the initial letter of key words.

Side headings indicate chapter sup-divisions if used with centre headings. They indicate major chapter divisions if used without centre headings.

Paragraph headings. Paragraph headings are used to indicate divisions within sub-divisions. Each paragraph heading is underlined and typed in lower case, and separated from the text above by three single spaces.

6. *Pagination*. Every page of the research report is given a number, although not every page has its page number typed on it. The title page or the initial page of the chapter, bibliography, or appendix does not have a page number typed on it, but a number is allowed for it in the series. There are two separate series of page numbers. The preliminaries are numbered using small or low case roman numerals (*i*, *ii*, *iii*, and so on). The title page is assigned the number *i*, but this number is not typed on it. Page numbers of the preliminaries should be placed in the centre at the bottom of the page, and are positioned one inch from the bottom of the page without punctuation. Roman numerals end with the last page or preliminaries preceding the main body of the report.

Arabic numerals are used to number all other pages beginning with the first page of chapter I. The bibliography, appendices, and pages of tables and figures are numbered serially and consecutively, following the last page of the last chapter. This number appears without punctuation in the top right-hand corner of the page. It is positioned one inch from the top and one inch in from the right-hand edge of the paper. The first page of chapter I is assigned number 1 although this number is not typed.

7. *Quotations.* When a researcher cannot rephrase a law, mathematical formula, or idea as concisely, accurately, or convincingly as the original author has phrased it, he uses the exact words of the author. Exactness means using the same words, the same punctuation, the same spelling, the same capitalization.

Short and long quotations. The basic form of a quotation is initially determined by its length. Quotations may be short or long, and the rules for presenting them differ. A short quotation usually occupies about four lines or less of the text. Such a quotation is incorporated into a sentence or paragraph frame work, without disrupting the flow of the text. It is double-spaced and enclosed in double quotation marks. A quotation that appears within a short quotation is enclosed in single quotation marks.

Example: Holizenger and Burt bifactor methods try to analyse a battery of tests in such a way that each test yields some relationship to a 'g' factor and to one, but only one, group factor. The group factors run through mutually exclusive groups of tests each, while the 'g' factor is common to all the test in the battery. The methods given by Burt and Holizenger "impose some arbitrary restrictions, one of which is requirement of a 'g' factor. A 'g' factor can always be found, if one insists upon it. But this theoretical bias may well predispose us to quite misleading results."

It is possible to acknowledge the above quotation in two ways. First, by placing a number superscript above the final word in the quotation:

"...this theoretical bias may well predispose us to quite misleading results."[1]

The source of quotation is then given in a footnote. Second, by using an abbreviated and more direct method of acknowledging the quotation: "...this theoretical bias may well predispose us to quite misleading results" (Guilford, 1954, p. 478).

In this example, instead of using a footnote to acknowledge the quotation, the author's name, the date of publication and the page number are presented directly after the quotation in the parentheses

The full reference of the source is given in the bibliography instead of footnote.

A long quotation usually consists of five or more lines of the text. It is single-spaced and indented three spaces from the left margin. No quotation marks at its beginning and end are used. Exceptionally long quotations may be placed in the appendix.

Example: Proper schooling is to be defined in terms of not merely physical and intellectual growth but of emotional and social development as well. For such developments, the principles of grouping and the effects of group membership on pupils in schools cannot be neglected. Just as in the family, relationships between children and parents are based on affection and discipline, and among equals are based on rivalry, competition or cooperation, so in the school there are social relationships for pupils between pupils and teachers, and between pupils and pupils. In this group:

> [The teacher, too, serves as stimulus for the learners. They are aware not only of the teacher's personal appearance but also of his speech and actions. If the teacher behaves in such a way as to be strongly admired by a pupil, the pupil will attempt to pattern his behaviour after the teacher. Similarly if the teacher attempts to meet the minor aggressive reactions of a pupil with aggressive acts against him, the pupil's behaviour in turn is likely to become more aggressive. Students attending the same school for a year or two expect to behave somewhat different in various classes because of the generally established reputation of the teacher (Klausmeier, 1961, pp. 119-120).]

Quotations of poetry. Short poetry quotations consisting of either a single line or part of a line should be enclosed in double quotation marks and run into text. In longer poetry quotations, quotation marks are not used. The lines are single spaced and indented. A slash is used to separate the lines (for example, "For many a false man as a friend/Came knocking all day at my door"). Between the verse, there is a double space.

Quotations in footnotes. When a quotation is placed as a footnote rather than in the text of the paper, the entire quotation is enclosed in quotation marks. It should be single spaced.

Ellipses. They are used to indicate omissions in quotations which are too long and not completely relevant. They can occur at the beginning or the end of a quotation. An ellipsis (...) is indicated by three full stops with the space before and after each full stop, in addition to between each full stop and the next. It should be used with utmost care so that the meaning of the original extract is not altered.

8. *Footnotes*. As the name implies, footnotes are normally found at the foot of a page. They serve a number of purposes. Footnotes enable the writer to validate and substantiate a point, statement or argument; to explain, supplement, or expand material that is included in the main body of the report; to provide cross references to materials appearing in other parts of the report; to acknowledge and give credit to sources of material that he has quoted directly or indirectly and to provide the reader with sufficient information to enable him to consult sources independently and thus to verify the authenticity and accuracy of material used. Footnotes should not be included for mere purposes of scholarly appearance. They should be used sparingly; only when the material being presented clearly needs acknowledgement or amplification.

Content and placement of footnotes. The footnotes usually include the source of information (usually the name of the author), the title of the source, exact page(s) of the source of reference, date of publication, and publisher and place of publication.

Reference to footnotes is made by the use of superscripts (i.e., numerals raised one half space below the dividing line) in the body of the text where the particular reference is given. The superscript is placed without any punctuation and without any space intervening between it and the preceding word or punctuation mark. The superscript may follow immediately the word or phrase to which the reference is made or it may be placed at the end of the sentence in which the reference has been made. With quotations the superscript is always placed at the end of the quotation.

Footnotes are always placed at the bottom of the page even if the context does not fill the page. They are separated from the text by a two inch horizontal line drawn from left margin, one double space below the last line of the text. The footnotes observe the regular paragraph indentions and are single spaced if they occupy more than one line. A double separates successive footnotes.

Example: In general, it seems a fair assumption, that teachers whom learners like best are those who have the best interaction with pupils and therefore the potentiality for greater success in teaching.[1]

---

1. [2]William M. Alexander and Paul M. Halverson, *Effective Teaching in Schools* (New York : Rine-hart and Co., Inc., 1957), p. 30.

If pupils like a teacher, they feel interested in the school and the subject he teaches them. Sudies[2] show that well liked teachers make well liked subject; that the subjects taught by well liked school teachers are voluntarily continued by students, while subjects taught by disliked teachers are continued only if they are required.

---

[2] C.W. Morse and G.W. Wingo, *Psychology and Teaching* Bombay : D.B. Tarapore vala Sons and Co., 1968, p. 234.

Some conventions in footnotes. 1. *Number of footnotes.* Footnotes are usually numbered consecutively within a chapter, or continuously within the entire report if chapter headings are not used.

2. *Citation of books.* The first footnote reference to each book should contain the full name of the author in its normal order (that is first name or initial and second initial precede surname); the title of the book cited; the place and date of publication in parentheses immediately after the title, followed by a comma; and the number of volumes and page in a series. The title of the book should be underlined and it should be given in full the first time it is cited. After the first-citation, the title may be abbreviated either by ellipses, by a shortened title, or by word abbreviation. If the researcher also wants to give the sections or chapter of the book, then all such should be inserted in quotation marks immediately after the author's name and should be followed by the title of the book underlined. A comma should usually be inserted after the title of a section or chapter of a book.

The abbreviations "vol." for volume, 'p.' for page and 'pp.' for pages should be used. The volume number should be written in Roman capitals, followed by a comma. The arabic numerals should be used in referring page numbers. If a book has two or three authors, the second and third authors' names are written in the normal order. If there are more than three authors, the name of the first is given, followed by "et al" or "and others". An editor who has collected the readings is indicated by placing the abbreviation, 'ed.' in the parantheses following the name.

Example: (*a*) *One author*

[1] Frederick L. Whitnery, *The Elements of Research* (New Jersey: Prentice-Hall Inc., 1937), p. 73.

(*b*) *Two authors*

[2]Halen M. Walker and Joseph Lev, *Statistical Inference* (Calcutta: Oxford and I.B.H. Publishing Co., 1965), pp. 71-75.

(*c*) *Three authors*

[3]Carter V. Good, A.S. Barr and Douglas E. Scates, *Methodology of Educational Research* (New York: Appleton Century Crofts, Inc., 1941) p. 17.

(*d*) *More than three authors*

[4]John R. Chandler and Others, *Successful Adjustment in College* (New York: Appleton Century Crofts, Inc., 1950), p. 72.

(*e*) *Editor of a collection of readings*

[5]Leon Festinger and Daniel Katz, (eds.), *Research Methods in Behavioural Sciences* (New York: Holt, Rinhart, and Winston, Inc., 1950), pp. 35-39.

(*f*) *No author given*

[6]Seven Ages (author not given) (New York, 1948), pp. 10-11.

Citation of articles. The first-note reference to each article should include (1) the author's name, with the first name or initials first; (2) the title of the article and the name of the magazine or journal; (3) the volume number; (4) the date of publication; and (5) the page number. The title of the article should follow the author's name. The title, followed by a comma, should be in quotation marks. The name of the magazine or journal should be underlined and it is followed by volume number. The date of publication should be put in parentheses and should be followed by a colon and page number. The abbreviations "vol." and "p.", should be omitted when both are given. If an article has two or more authors, the names of authors are cited in the same manner as in case of books.

Examples: [1]Louis W. Rapeer, "The Problems of Formal Grammar in Elementary Education", *Journal of Educational Psychology*, 4 (March, 1913), 125-137.

[2]F.W. Warburton, *et. al.*, "Predicting Student Performance in a University Department of Education", *British Journal of Educational Psychology*, 33 (March, 1963), 68-79.

Use of "Ibid" "Op. cit." and Loc. cit." When the same work is referred to more than once in a footnote, the reference is not repeated in its full form. The abbreviations *Ibid.*, *Op. cit* and *Loc. cit.* are variously used to indicate the source. When a work referred to once is referred again and no other reference occurs in between, the footnote includes the name of the author followed by a comma, the abbreviation "ibid" underlined followed by a comma and the page number.

Example: [1]J.P. Guilford, *Fundamental Statistics in Psychology and Education* (New York: McGraw Hill Book Co., 1956), p. 274.

[2]*Ibid.* (indicates same page as previous reference).

[3]*Ibid.*, pp. 287-290.

When a work which has been referred to once in the footnote is acknowledged again within a page or two after the occurrence of other references in between, the footnote includes the surname of the author followed by a comma, and the abbreviation *Op. cit.* underlined followed by a comma and the page number.

Example: [1]David White, *Statistics for Education with Data Processing* (New York: Harper and Row, 1971), p. 189.

[2]Henry E. Garrett, *Statistics in Psychology and Education* (New York: Longmans, Green and Co, 1958), p. 277.

[3]White, *Op. cit.*, p. 189.

When a second but non-consecutive reference follows, referring to the same work and same page previously cited, *Loc. cit.* underlined is used instead of *ibid* and the page number is not repeated.

Example: [4]Garrett, *Loc. cit.* (This reference is to page 277).

After a reference has been made to two or more works of the same author, *Op. cit.* and *Loc. cit.* may not be used to refer any of these works. In such cases it is desirable for the sake of clarity to repeat full title or to use short title of the work than to use these abbreviations.

9. *Construction of tables.* Tables are used to convey information. They are used to help readers spot important details, see relationships, get a brief overview of the findings, or grasp the significance of data much more quickly and conveniently than through a long textual discussion. Although the use of the tables is not recommended simply to repeat information adequately covered in the text of the report, the text should contain sufficient detail to support the particular argument being put forward. Some readers feel scared of the figures in the tables and prefer to read the evidence presented in the main text. Other readers find tabular presentation easier to follow than written discription.

Content and placement of tables. The essential characteristics of a good table are simplicity and unity. There is no use in including too much data in a single table. It is advisable to concentrate on a limited number of ideas in a single table and to use several tables for making several comparisons of distinctly different kinds of data. If several comparisons of distinctly different kinds are combined in one table, or if like comparisons are separated into many different tables, a reader may get lost and confused. A well constructed table integrates and presents several related facts in one main idea.

Tables should always be introduced by a suitable heading. They should always follow as closely as possible to its first mention in the text. For example, a table may be introduced in the following way:

Means and Standard Deviation of Boys and Girls are given in Table 11.2.

if a table is large enough to cover more than a half-page, it should be placed and centred on a page by itself. If it is short and occupies less than half a page, it may be placed on the page with textual material, preferably following as closely as possible the textual discussion that relates to it. In practice, it is often more convenient to have all tables on separate pages. This way they will not have to be recopied each time the report is revised. When the final draftof the report is prepared, the original tables be inserted be in the proper places.

Long and detailed tables that interrupt the continuity of the texual discussion may be put in the appendix. Here the researcher must be guided by the utility of the material being presented in the table. If the material is essential for the conclusions which follow, it should be presented in the main body of the text. If, however, the material consists of supporting details to which the interested reader may refer to, if he wishes, then its proper place is in an appendix.

Numbers and captions. All tables including those that appear in the appendix should be numbered to permit their easy identification by the reader. The usual practice is to use arabic numerals and to number tables consequently throughout the report. If tables are included in the appendix, the numbering sequence should continue from the last table number in the text. The word 'table' in capitals, followed by its number, is placed alone on the first line and the heading or title two spaces below it. Both table and its number are typed in full capital letters and no end punctuation is necessary.

Every table has a title or a heading. It is a concise summary of what is presented in the table. The title should be self-contained and not a complete sentence. It is placed two spaces below the word

'Table' with no terminal punctuation, if the title is more than one line, it is single spaced and centred like an inverted pyramid. Phrases like "Table Showing", "Distribution of", or "Frequency of" (for instance, Table showing Average Income of Parents of Scheduled Caste Students of Dayanand High School, Simla for 1978) should be avoided. The title of each table must correspond exactly with that given in the List of Tables.

The top of the table is placed three spaces below the last line of the title. Column headings, of box heads, should be worded carefully and clearly, describing the nature and units of measure of the data listed. When a table is divided into sections and rules are drawn to separate sections, every column heading shold be centred above the column. The column headings should be parallel in grammatical structure making use of some common abbreviations. Such abbreviations as No., % and *f* may be used for such terms as number, per cent and frequency. The percentage symbol (%) should be placed at the top of the column not with the numbers in the table. In some cases, the column heading may be so long that it will be necessary to turn the paper broadside and to type from the bottom of the page.

To facilitate locating items referred to in the table, the columns of a table may be placed in parentheses and centred below the column heading. The first column is numbered (1) and the other columns are numbered consecutively.

The zero preceding a decimal, the degree sign, and the rupee sign are usually omitted from all entries except the first and the last. When the column is carried over to another page or there is a break in the column, the zero, degree sign, or rupee sign must be repeated in the first entry. When no data are available for a particular cell, a dash rather a zero may be put in the particular cell.

Rulings and size. Rulings or lines are used only if they facilitate the reading of the table. A double or heavy horizontal line, separating the table from the title of the table, is placed above the column headings or captions and a single line below them. A horizontal line is also placed at the bottom to separate the table from the textual material which follows three space below. Vertical lines and additional horizontal ones may be added if they break up the data into logical and simpler groups. Vertical lines at the left and right margins of the table are not used.

Tables should be no longer than the pages of the report. Whenever possible a table should be completed on a single page. Tables larger than page size should not be folded into the report. If a table will not fit into the normal page, lengthwise or crosswise, it may be reduced in size using a photographic, process or using small print.

When such a photographic reduction of long tables is impracticable, the tables may be continued from page to page. In such a situation, the table is not closed until the final page. The word "table" and the "number" are repeated, followed by the word "continued", at the top of the succeeding pages. (For example, Table 5—continued). All other column headings or captions except the title are repeated. Wide tables may be divided and placed on two pages facing each other. In an instance of this kind, the full heading is placed on the top of the first page and only the word "table" and the number, followed by the word "continued" appear on top of the opposite page.

Footnotes. Footnotes are also used to explain certain items in the table. The footnotes for a ruled table should be provided immediately below the horizontal line that closes the table. The footnotes for an unruled table should appear immediately below the last entry in the column. Numerical superscripts should not be used to indicate the footnotes as they would be confused with the data contained in the table. Small letters (a, b, c) or certain typewriter key symbols are appropriate to denote footnotes. A single asterisk (*) may be used to indicate one footnote, and additional asterisks may be used to denote additional footnotes. If there are more than three footnotes, small superscript lower case letters (a, b, c, d,) instead of asterisks may be used. The first footnote is placed one double space below the table and if it occupies more than one line, the footnote is single spaced. There is a double space between footnotes.

10. *Figures.* Figures are devices which are used by a researcher for the purpose of presenting data clearly and concisely. When skillfully used, they reveal important trends or relationships that a reader might not grasp when examining complex statistical data. Figures do not replace textual description, but they may help a researcher to explain and interpret complicated data effectively to the reader.

Format and placement. The format of figures should be as simple as the data presented permit and should convey the exact idea without the aid of textual description. Complicated or carelessly prepared figures are less effective communicators than words. Figures should be used sparingly, and used only if they snap significant ideas or relationships into a sharp focus for the reader more quickly and effectively than other means of presentation.

difficulties in reading, large figures may be reduced in size by photocopying. It is a good technique to draw large figures on cartridge paper for subsequent photographic reduction since this smooths irregularities and helps to give a polished presentation.

Number ana captions. Figures are labelled with the word "figure" or "fig" and are numbered consecutively throughout the report to permit their easy identification by the reader. They should be numbered with arabic numerals and should be labelled to specify exactly what the presentation includes. The word figures and its number should be typed one double-space below the figure. The caption or title of the figure may be placed after the figure number rather than above the figure. It may be typed or lettered in an inverted pyramid form with no end punctuation or in paragraph, underhung, or block form. After adopting any of the title forms, the researcher must follow the same pattern throughout the report.

Footnotes. Footnotes are often used, as in the case of tables, to explain a point within the figure or to define an abbreviation or symbol. Letters of alphabet or special symbols such as *, †, etc. are, used for referencing rather than arabic numerals. The spacing rules for footnotes as discussed in Tables are equally appropriate when applied to figures.

Types of figures. The commonest types of figures used in reports are line graphs, bar graphs, pie charts, maps, organization charts, diagrams of apparatus and photographs. Campbell (1969), Dugdale (1967) Turabian (1967) and Best (1977) have explained in detail the methods of presenting theses figures, the merits and demerits of using each type, and the rules governing their construction.

The line graph. The line graph is used in the form of the frequency polygon, the cumulative frequency graph, or the cumulative percentage curve or ogive. This graph is particularly useful in presenting relationships between different types of data, and in making predictions based upon trends.

A pair of co-ordinate axes are usually used to represent independent variable and measured characteristics. The horizontal axis (x-axis) measures the independent variable and the vertical axis (y-axis) the measured characteristic. Graphical arrangement should proceed from left to right on the horizontal axis and from bottom to top on the vertical. The point of intersection (origin) of the axes should be represented by the zero point and scale intervals should be equal.

When the graph paper is used, the co-ordinate axes should be made

Example:

TABLE 1

Significance of the Difference between Mean Scores of Popular and Unpopulars on Different Sprangerian Values*

Value	N=100 *Popular teachers*			N=100 *Unpopular teachers*					
	Mean	*Standard deviation*	*Standard error of mean*	*Mean*	*Standard deviation*	*Standard error of mean*	*Mean difference*	*Standard error of mean difference*	*t*
1	2	3	4	5	6	7	8	9	10
Theoretical	41.62	7.43	0.552	34.59	7.98	0.637	7.03	1.090	6.45**
Economic	41.98	7.30	.533	47.07	8.62	.743	5.09	1.129	4.51**
Aesthetic	32.29	7.43	.552	46.46	11.47	1.316	14.17	1.366	10.37**
Social	44.01	6.47	.419	37.95	7.43	.552	6.06	.985	6.15**
Political	41.62	5.53	.306	37.69	5.30	.281	3.93	.766	5.13**
Religious	38.48	8.29	0.687	36.24	6.74	0.454	2.24	1.068	2.09***

*Adapted from L. Koul, *Factorial Study of Certain Personality Variables of Popular Teachers in Secondary Schools*. (Unpublished Doctoral Thesis, Kurukshetra: Kurukshetra University, 1972) p.94

**Statistically significant, p. <.01

***Statistically significant, p. <.05

Figures should be accompanied by appropriate tables to represent the numerical data they are based on, if the data are not included in the figure itself. A figure should not be introduced into a report without a discussion of it. There should be a triple space between the textual introduction and the figure.

Figures that are smaller and occupy less than a half page should be placed on the same page as textual discussion. Those that occupy more than a half page should be placed on a separate page. If photographic facilities are available and if reduction does not cause heavy enough to distinguish them from the other lines. When several lines representing different data are drawn, they may be distinguished by using ink of different colours or by using different types of lines, namely solid, cotted, or alternate dots, and dashes.

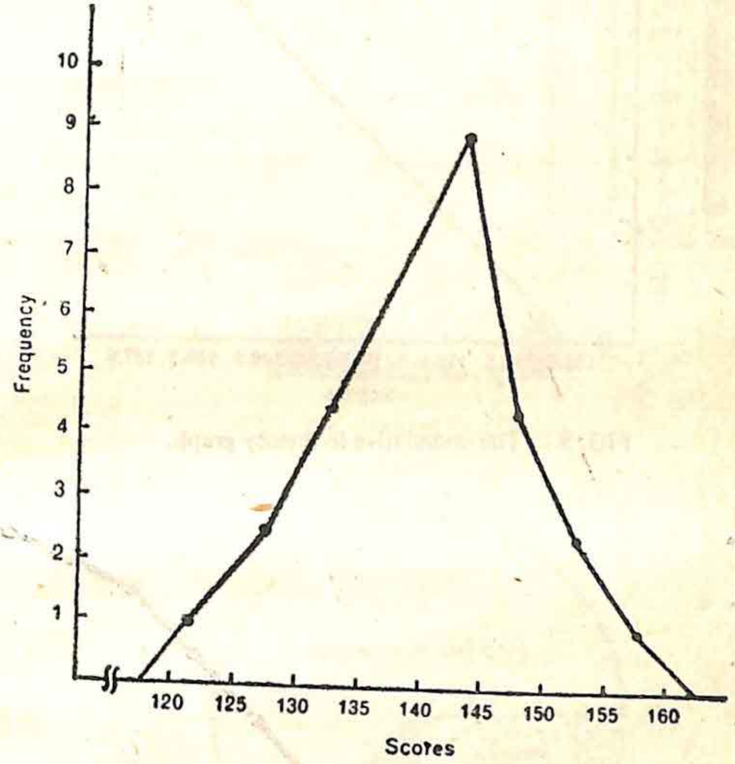


FIG. 9.1. The Frequency Polygon.

The bar graph or diagram. The bar graph or diagram is used to compare the relative amounts of some trait or attribute, like intelligence, height, weight or academic achievement, possessed by two or more groups. It is also used to compare several different variables. Examples are teachers' salaries by districts, relative expenditure for school, college and university education, etc. The bar graph may be used in horizontal, vertical or divided form, representing data by bars of equal width, drawn to scale-length with numerical data lettered within the bar or outside it.

The pie diagram. The pie diagram is useful when the researcher wishes to picture proportions of the total in a striking way. For example, a researcher may use it to explain the distribution by caste of pupils in eighth grade of a school in the city of Chandigarh.

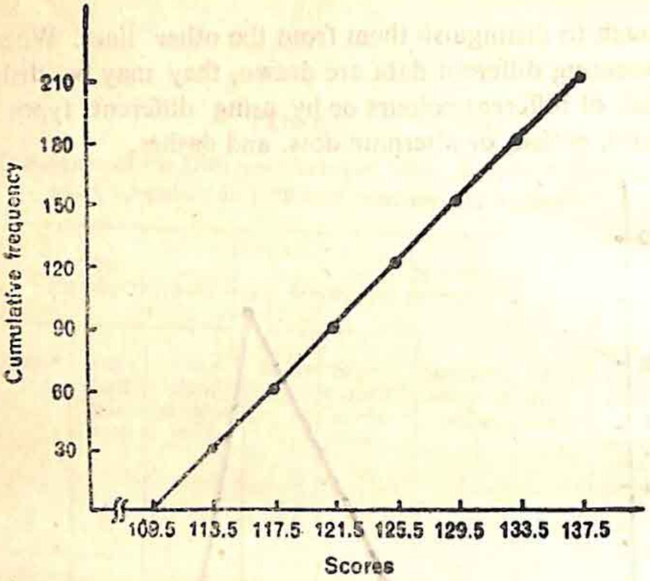


FIG. 9.2. The cumulative frequency graph.

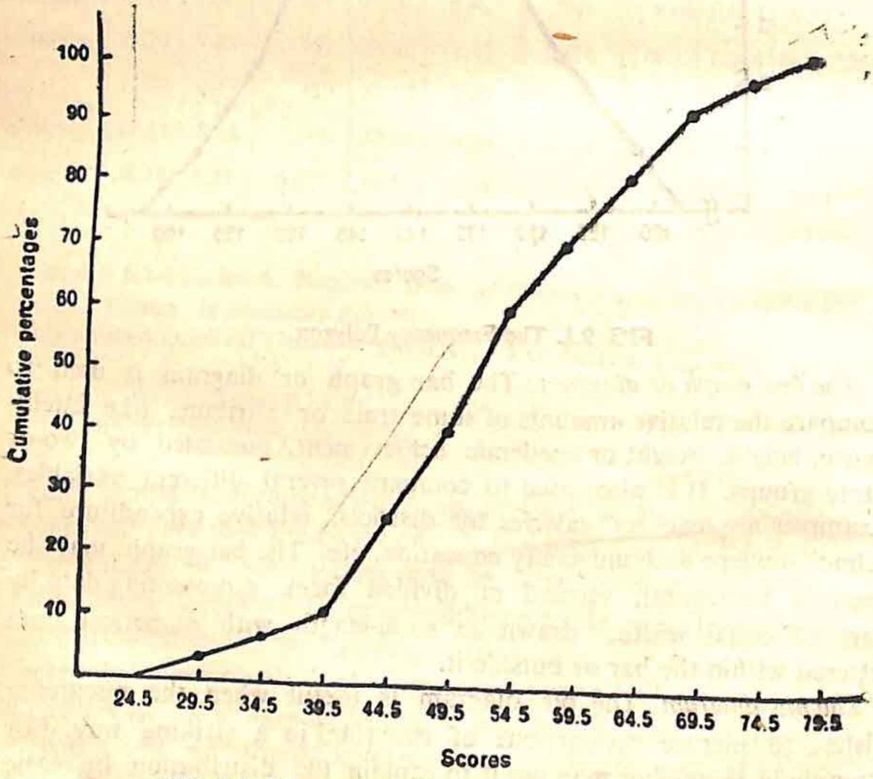


FIG. 9.3. Cumulative percentage curve or ogive.

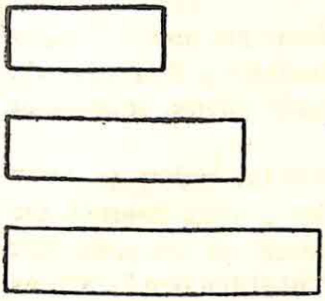


FIG. 9.4. Horizontal Bar Graph.

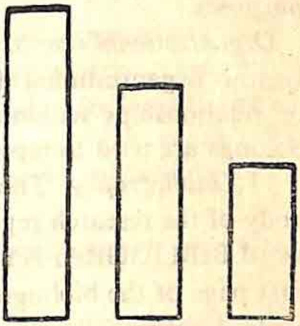


FIG. 9.5. Vertical Bar Graph.

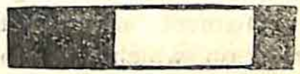
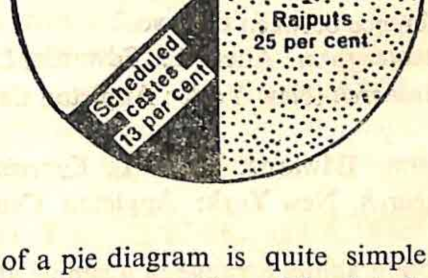


FIG. 9.6. Divided Bar Graph.



The construction of a pie diagram is quite simple. There are 360 degrees in the circle. The proportion of the data is indicated by the number of degrees in each segment of the 360° circle. For example, 60 per cent of a data is indicated by 216° (60 per cent of 360°). The radius is first drawn vertically and then components are arranged proportionately in clockwise direction in descending order of magnitude.

Maps. Maps are used for geographic location or identifications. Dots, circles, or other symbols are used for identification within a map. Shading, cross hatching or colouring are also used for such purposes.

Organizational charts. The organizational charts are used to show various organizational functions like lines of authority, flow of work, or relationships within an organization. Usually circles, squares, or oblongs are used to represent various units.

11. *Bibliography.* The bibliography is typed at the end of the main body of the research report. It is preceded by a page bearing the word BIBLIOGRAPHY, capitalised and centred on the page. The first page of the bibliography has the word "BIBLIOGRAPHY" as centre heading.

The purpose of a bibliography is quite different from that of a footnote. The purpose of a footnote is to give the specific location of the source of the statement made in the text, including the number of the actual page on which the statement appears in the original source. The purpose of the bibliography, on the other hand, is to identify the whole work rather than a specific part of it. It would be inconvenient to repeat the general reference on each occasion that a source is cited in a footnote, yet the full details of the source need to be included somewhere in the report for the information of the reader. The style of citation and punctuation in a bibliography is quite different from a footnote. For example, in a footnote the author's name is given in the natural order of initials or first names followed by the surname, but in a bibliography the surname precedes the initials of first name. The following example illustrates the difference between the two.

One author-footnote form. Allen L. Edwards, *Experimental Design in psychological Research* (New York : Appleton Century Crofts, Inc., 1959), 112.

Bibliography form. Edwards, Allen L. *Experimental Design in Psychological Research.* New York: Appleton Century Crofts, Inc., 1959. p. 112.

Most of the bibliographies arrange in a simple alphabetical list the references used by the researcher in preparing the report, but some of them classify references under different headings such as books, journals, reports, newspapers, public documents and miscellaneous. The former method is preferred and generally used.

Style of citing. Each reference begins at the left hand margin of the page and material for each reference is single-spaced. There is double space between references. The references are usually listed alphabetically by author's surname. The surname is typed flush with the left margin so that it stands out clearly. Second and subsequent lines of the same entry are single spaced starting three spaces from the left margin.

There is no universally accepted style that may be used for constructing a bibliography. Variations in practice are often observed in citing references to books, journals, documents, newspapers, etc. However, there are some accepted styles for citing such references which may be helpful to a researcher to maintain consistency throughout the report.

(*a*) *For a book.* Each book reference in a bibliography is set up alphabetically in the following order:

(*a*) *For a single author.* Name of the author—surname first followed by a comma and the full first and middle names, followed by a comma.

For joint authors. Name of the authors—the name of the first author (surname first), followed by the full name of the other author or authors, first and middle names before the surname, with commas in between the different author's names, followed by a comma in the end.

(*b*) Title of the book (underlined), followed by a full stop.

(*c*) Place of publication, followed by a colon.

(*d*) Name of publisher, followed by a comma.

(*e*) Year of publication, followed by a full-stop.

(*f*) Number of page or pages, followed by a full stop (the abbreviation p. is used for a single page and pp. for more than one page).

Example. Ausubel, D.P., *Theory and Problems of Adolescent Development*. New York : Grune and Stratton, Inc., 1954, p, 111.

Barr, A.S., W.H. Burton, and L.S. Brueckner, *Supervision.* New York : D. Appleton Century Co., 1938, pp. 12-14.

If the bibliography is annotated, the annotation should begin a double space below the last line of the reference date and should be single spaced.

Example. Ary, D., Lucy, C. Jacobs, and A. Razavieh, *Introduction to Research in Education.* New York: Holt Rinehart and Winston, Inc., 1972.

Familiarizes the beginning researcher with the procedures for conducting an original research project. It focuses on the typical and practical problems encountered in research, beginning with the formulation of problem and continuing through the preparation of the final report.

(*b*) *For an article.* In case of journal article, the title of the article is enclosed in quotation marks. It is followed by a comma and the name of the journal underlined. The volume number in roman numerals should follow it. The place of publication and publisher are not included. The date of publication enclosed in parantheses follows the volume number. It is followed by a comma and page number.

Example. Koul, L., "A Study of the Impact of Teacher Training upon Attitude of Student-Teachers Towards Teaching", *Indian Educational Review*, XII (January, 1977), 38-43.

If the author or one of the authors is a woman, her first name is given in full.

Example. Vishnoi, Kusum, "Interest Patterns of High and Low Achievers: A Comparative Study", *Indian Educational Review*, XII (January, 1977), 44-48.

Chronological order. When a number of references by the same author are listed, each entry should be ordered chronologically.

Guilford, J.P., *Fundamental Statistics in Psychology and Education*. New York : McGraw Hill Book Company, Inc., 1950.

Guilford, J.P., *Psychometric Methods*. New York : McGraw Hill Book Company, Inc., 1954.

Guilford, J.P., *General Psychology*. Princeton, New Jersey : D. Van. Nostrand Company, Inc., 1964.

(*c*) *Some special cases*. References to encyclopaedias, public documents, newspapers, publications of learned organizations, reports, unpublished materials and edited works follows special usages slightly different from those observed for books and journals. The following examples will illustrate the variations.

Encycopaedia Articles. Bertocci, Peter A., "Personality" *The Encyclopaedia of Psychology*, edited by Philip L. Harriman, Philosophic Library, New York, 1946.

Reports. Government of India, Ministry of Education, *Educational and National Development* Report of the Education Commission, 1964-66, New Delhi.

Unpublished Materials. Bansal, J.P., "A Study of Need Differences Among Urban High School Boys and Girls at Different Levels of General Intelligence And Socio-Economic Status" Unpublished Doctoral Thesis, Department of Education, Kurukshetra University, Kurukshetra, 1975.

Miller, Lee W., Influence of the New Testament of Emerson's Teaching Method", unpublished paper read before the South-Central Modern Language Association, Biloxi, Mississipi, November 6, 1954.

Edited Works. Buros, Oscar, K., ed. *The Sixth Mental Measurement Year Book*. Highland Park, New Jersey: Gryphon Press, 1965, p. 1163.

Public Documents. Walker, Stanley, "Our Trivial Press" *1936 Essay Annual*, edited by Erick A. Walter, New York: Scott, Foresman and Company, 1936.

Newspaper Articles. Chabra, Rami, "Mainly a Matter of Women", *Indian Express*, November 3, 1978, p.5.

Editorial in *The Hindustan Times*, July 13, 1978.

Publications of Learned Organizations. National Society for the Study of Education, *Modern Philosophies of Education*. Fifty-fourth Year Book, Part I., Chicago: The University of Chicago Press, 1955, p. 374. Johnson, Wendell, "Teaching Children with Speech Handicaps."

The Education, of Exceptional Children. Forty-ninth Year Book, The Study of Education, 1950.

12. *Appendix*. The appendix is also preceded by a separate page bearing the word "APPENDIX" typed just above its centre in capital letters.

All the different items included in the appendix are grouped and labelled as Appendix A, Appendix B, etc., and listed in the table of contents. These headings are typed in capital letters at the centre of the top of the first page of each group of items. If there are no different groups of items, the first page of appendix should bear the capitalised general heading "APPENDIX".

13. *Index*. The index is preceded by a page containing the word 'INDEX', capitalised and centred on the page. The first page of the index has also the centred heading 'INDEX'.

The index items are arranged alphabetically by their key words. The index is set up in two columns. It is single spaced, with principal items beginning flush with the left hand margin, the carry-overs indented three spaces, and sub-items indented two spaces. The items are followed by a comma and the page number (s) where the items in the text can be found.

PROOF READING OF THE FINAL DRAFT OF THE RESEARCH REPORT

When the final typed copy of the report is completed, it is necessary to proof read each page carefully. The typing errors are best detected by proof reading each line from right to left. A few erasures may be made to correct a copy, but crossing out or inserting words or sentences and typing over letters are not permissible. The corrections should always be made by typewriter and not in ink. It is advisable to retype the page if there are more than one or two corrections. While retyping care should be taken to accommodate the material in such a way so that the last line on the page comes even. If no care is taken, the researcher may have to retype the rest of the chapter and renumber the remaining pages of the research report.

Mathematical and some other special symbols are not available on the typewriter. All such symbols should be inserted in India ink. It is necessary to check that the page numbers in the 'table of contents', 'list of tables' and 'list of figures' correspond exactly with 'chapter headings', 'table headings' and 'figure headings' in the text. Before stitching the pages together or binding the research report, a final check should be made that the pages are arranged in the correct sequence.

POST TEST

1. Describe in detail the general format of the research report.
2. Why is reporting of research study important?
3. What are the essential guidelines for presenting a research study?
4. Describe the salient features of a good research report.
5. Describe the style and format of writing the research report logically and objectively.
6. Describe in deatail the procedure which is followed in typing different sections of the research report.
7. Describe the uses of the following in the research report:
   (*a*) footnotes,
   (*b*) bibliography, and
   (*c*) appendix.
   How should they be presented in a typed research report?
8. Write down the standard abbreviations of some words which are generally used in the footnotes, the tables and the bibliography of research report.
9. Describe the procedure which is adopted in proof reading of the final draft of the research report.

10. Methods of Educational Research

The organization of this chapter is as follows. The first section deals briefly with the importance of research methods and their classification into three basic categories: historical method, descriptive method and experimental method. The second section describes in detail the historical method with emphasis on the following aspects: nature of historical research, value of historical research in education, types of historical research, and steps in historical research.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Describe briefly the importance of research method in the research process.
2. Name the three basic categories of research methods.
3. Define the term history.
4. Describe the nature of historical research.
5. Differentiate between history and historical research.
6. Describe the value of historical research in education.
7. List and describe the types of historical research.
8. List and describe the steps which the researcher may adopt in conducting the historical type of research.
9. Explain the meanings of the following:
    - (*a*) Sources of data: primary sources and secondary sources of data.
    - (*b*) Criticisms of data: external criticism and internal criticism of data.

In the foregoing chapters we have been concerned with the meaning and areas of educational research, and the various stages that the researcher has to undergo while planning and conducting a research study. The researcher first selects the area of research; he identifies and defines his problem; he reviews the literature to acquaint himself with the recent developments in knowledge related to his problem; he states hypotheses and defines specifically the objectives of his study; he finalizes his research plan; and then starts working actively on the problem itself according to the plan. In the planning process, the researcher has to decide about research method that he could use in solving his research problem.

Research methods are of utmost importance in a research process.

They describe the various steps of the plan of attack to be adopted in solving a research problem, such as the manner in which the problems are formulated, the definition of terms, the choice of subjects for investigation, the validation of data-gathering tools, the collection, analysis and interpretation of data, and the processes of inferences and generalizations.

Methods of research, according to Good, Barr and Scates (1941 p. 207), may be classified from many points of view; the field to which applied: education, history, philosophy, psychology, biology, etc.; purpose: description, prediction, determination of causes, determination of status, etc.; place where it is conducted: in the field or in the laboratory; application: pure research or applied research; data gathering devices employed: tests, rating scales, questionnaires, etc.; nature of the data collected; objective, subjective, quantitative, qualitative, etc.; symbols employed in recording, describing, or treating results: mathematical symbols or language symbols; forms of thinking: deductive, inductive etc.; control of factors: controlled and uncontrolled experimentation; *methods employed in establishing casual relationships*: agreeme[illegible] difference, residues and concomitant variation.

Research methods when cl[illegible] in the above manner give rise to a comprehensive list with mu[illegible] of overlapping among the different methods. Yet despite this lack of clear-cut distinction among the methods, most authors agree on three basic categories:

1. *Historical method*: Which provides a method of investigation to discover, describe and interpret what existed in the past.
2. *Descriptive method*: Which provides a method of investigation to study, describe and interpret what exists at present.
3. *Experimental method*: Which provides a method of investigation to derive basic relationships among phenomena under controlled, conditions or, more simply, to identify the conditions underlying the occurrence of a given phenomenon.

The discussion of research methods in the present text will also be organized according to the basic categories outlined above.

The selection of a method and the specific design within that method appropriate in investigating a research problem will depend upon the nature of the problem and upon the kind of data that the problem entail. However, the method selected should be in harmony with scientific principles and adequate enough to lead to dependable generalization. In any specific study, although it is a common practice to use any one of the above methods, yet there is no reason why two or more methods cannot be applied effectively in combination in certain research situations. For example a researcher may seek the solution of a problem by studying its history through an examination of documents and then determining, its present status by some sort of survey.

A researcher must have a thorough understanding of all research methods with particular reference to their strengths, limitations, applicability, and appropriateness. It will help him to carefully plan the steps he will take in the research process and describe the method clearly before he actually starts working on the solution of the problem. According to Hillway (1964, p. 138):

> . . to describe in detail the specific method being used, incidently, constitutes a very good way of determining whether the method chosen has been worked out properly and is likely to prove effective. If the scholar cannot describe his method, the chances are that it is too vague and general to yield him satisfactory results.

A pre-planned and well-described method will provide the researcher a scientific and feasible plan for attacking and solving the problem under investigation.

THE HISTORICAL METHOD

History is a meaningful and an organised record of past events. It is not merely a list of events arranged chronologically, but a valid integrated account of social, cultural, economic and political forces that had operated simultaneously to produce a historical event. Education too, has a history of its own which has progressed under the influence of such forces. An understanding of the historical background of education would enable the educator to recognise the ills of most educational practices which were tried in the past and found wanting.

NATURE OF HISTORICAL RESEARCH

Historical research attempts to establish facts so as to arrive at conclusions concerning past events. This is usually accompanied by an interpretation of these events and of their relevance to present circumstances and what might happen in the future. The main purpose of historical research, therefore, is to arrive at an accurate account of the past so as to gain a clearer perspective of the present. This knowledge enables us at least partially to predict and control our future existence.

Historical research, as any other type of research, includes the delimitation of a problem, formulating hypotheses or tentative generalizations, gathering and analysing data, and arriving at conclusions or generalizations based upon deductive-inductive reasoning. However, the historian faces greater difficulties than researchers in any other field. He, according to a Ary *et al* (1972, p. 283), lacks control over both treatment and measurement of data, has relatively little control over sampling, and has no opportunity for replication. Historical data is a closed class of data located along a fixed temporal locus and historian has no choice of sampling his data, and he is supposed to include every type of data that comes his way. Historical research is not based upon experimentation, but upon reports of observation which cannot be repeated. The historian handles a data of unique type. They are mainly traces of past events in the form of various types of documents, relics, records, and artifacts having a direct or indirect impact on the event under study.

The job of the historian becomes more complicated when he derives truth from historical evidence. The major difficulty lies in the fact that the data on which historical research is based are invariably relatively inadequate and at times the study is conducted with all of the independability that the data may entail. According to Best (1977, p. 344):

> The historian must depend upon the reported observations of others, often witnesses of doubtful competence and sometimes of doubtful objectivity.

These obviously poses a difficulty in matters of objectivity of interpretation.

The date of occurrence of a certain historical event is another difficulty. It may be difficult to determine it partly because of changes brought out in the system of calender and partly due to incomplete information.

VALUE OF HISTORICAL RESEARCH IN EDUCATION

Historical research has great value in the field of educational research because it is necessary to know and understand educational achievements and trends of the past in order to gain perspective on present and future directions. Knight (1934), as quoted by Good, Barr and Scates (1941, p. 243) has given the following analysis of the value of historical research:

1. A knowledge of the history of schools and other educational agencies is an

important part of the professional training of the teacher or the school administrator.

2. Much of the work of the school is traditional. The nature of the work of the teacher and the school administrator is restrictive and tends to foster prejudices in favour of familiar methods. The history of education is the "sovereign solvent" of educational prejudices.
3. The history of education enables the educational worker to detect fads and frills in whatever form they may appear, and it serves as a necessary preliminary to educational reform.
4. Only in the light of their origin and growth can the numerous educational problems of the present be viewed sympathetically and without bias by the teacher, the school administrator, or the public.
5. The history of education shows how the functions of social institutions shift and how the support and control of education have changed from very simple and local arrangements to those that are now somewhat centralised and complex.
6. The history of education is an ally in the scientific study of education rather than a competition. It serves to present the educational ideals and standards of other times, and it enables social workers to avoid mistakes of the past.
7. It inspires respect for sound scholarship and reverence for great teachers.

TYPES OF HISTORICAL RESEARCH

Education has a history that needs to be studied in scholarly detail. Historical studies that could be conducted with profit to the field of education may include the following:

1. Bibliographic research.
2. Legal research.
3. Studying the history of ideas.
4. Studying the history of institutions and organizations.

Bibliographic Research

Bibliographic research aims at determining and presenting truthfully the important facts about the life, character, and achievements of important educators. In Indian context one may study the contributions of Gandhiji, Tagore and other leading educationists and their influence on current educational practice and thought.

Legal Research

Legal research is of immense value and interest to educational administrators. It aims to study the legal basis of educational institutions run by different religions and castes, relation between central and state governments with regard to education, legal status of teachers and students, administration of private aided schools, school finance, participation of students in the administration of universities, etc. Legal research needs special training in the field of law, and any one without this training is not competent to do this type of research.

Studying the History of Ideas

Studying the history of ideas involves the tracing of major philosophical or scientific thoughts from their origins through their different stages of development. It also aims at tracing of changes in popular thoughts and attitudes over a given period of time. The evolution of current concepts like team teaching, the problem-solving approach, mastery-learning approach, etc. provide important topics for historical research.

Studying the History of Institutions and Organizations

Studying the history of some prominent schools, universities and other educational institutions also provide numerous problems for significant historical research. When studying such history, the same general method applies as for the study of an educator's life. In India for example, one may study the history of the growth and development of Vishwa Bharti University.

STERS IN HISTORICAL RESEARCH

The steps involved in undertaking a historical research are not different from other forms of research, but the nature of the subject matter presents a researcher some peculiar problems and requires him to apply some special standards and techniques. In general, historical research involves the following steps:

1. Selection of the problem
2. Formulation of hypotheses
3. Collection of data
4. Criticism of data
5. Interpretation and reporting of findings.

Selection of the Problem

A researcher may select a problem pertaining to the history of individuals, institutions, organizations, law, curriculum, administration, textbooks, teacher education, equipment, important concepts and thoughts that have influenced education during a specific period of time in a given culture or sub-culture determined by religion, caste, sex, age, or work. He may delimit his study to an era of events in a local, regional, or national setting, or he may study the trend of events in different eras, different societies, or different cultures. The historian may discover new knowledge, the meaning of which, when interpreted will provide answers about past events. Sometimes he may doubt an old interpretation of existing data and then attempt to provide a more satisfactory explanation of past events.

The researcher should exercise due care in selecting and delimiting the historical problem for investigation. He should check that the problem selected should not only be of historical and current significance, but answerable by available methods of research and by the available sources of data. Sometimes many worthwhile topics of historical importance may have to be discarded when adequate data are not available.

Formulation of Hypotheses

The hypotheses that the researcher constructs for historical research are useful in explaining events, conditions or phenomena of the historical period in question. Sometimes it is argued that in such type of studies a researcher is merely interested in concrete events in their singularity, he has merely to check the validity and authenticity of facts about past events and arrange them in a chronological sequence. Therefore, the researcher may not formulate any hypotheses in such investigations. But the findings based on unstated hypotheses are ambiguous and do not explain or describe the structural interrelations of the phenomena under study. The reports of such findings relate what happened in the past but do not explain how and why the events occurred in a particular sequence.

However, it must be noted that the hypotheses for historical research may not be formal hypotheses to be tested. Rather, they are written as explicit statements that tentatively explain the occurrence of events and conditions. Van Dalen (1973, p. 177) is of the opinion that:

> An explicit statement, or a systematic awareness of the hypotheses used in determining what facts were significant, clarifies the relationship between the investigator's ideas and the facts he reports, and it minimizes the possibility of employing trivial, biased, conflicting, faulty, or archaic hypotheses.

While formulating hypotheses, a researcher may formulate questions that are most appropriate for the past events he is investigating and then directs his research towards seeking answers to these questions with the help of evidences.

Collection of Data

After the problem has been selected and stated and appropriate hypotheses or questions have been formulated, the researcher has to collect all the data available so that hypotheses may be thoroughly verified. The collection of data in historical research is a tedious and time consuming process. The researcher usually sifts through the vast materials of human activity that testify about past events, and from these he identifies and selects data that are relevant to his problem. These data are classified into primary and secondary sources. It is important for a researcher to distinguish between them and develop skill in locating them.

Primary Sources

Primary sources are eye witness accounts and are the only solid bases of historical enquiry. Good, Barr and Scates (1941, p. 253) have called them as the "first witnesses to a fact". The original documents or remains come under the category of primary sources. They are available in written, pictorial and mechanical forms as under:

1. *Personal records*. Certificates, diaries, autobiographies, affidavits, declarations, letters, wills, deeds, contracts, and original drafts of speeches, articles, books, and pamphlets

2. *Official records*. Legislative, judicial, or executive documents prepared by central or state governments, municipalities, panchayats or other local bodies, such as constitutions, laws, charters, court proceedings and decisions; the data preserved by missionaries and other religious organizations such as financial records and records of the minutes of the meetings of managing or governing bodies; the information compiled by central or state education departments, special commissions, professional organizations, school boards, administrative authorities, such as the minutes of meetings, reports of committees and commissions, administrative orders, school surveys, annual reports, budget, attendance records, cumulative records of dramas, games, musical and athletic events, and examination

3. *Oral testimony of traditions and events*. Myths, folk tales, family stories, ceremonies, spoken account of a witness of an event, interviews with administrators, teachers, students, parents or gaurdians, school patrons, and prominent educationists.

4. *Pictorial records*. Photographs, movies, micro-films, drawings, paintings, coins, and sculpture.

5. *Mechanical records*. Photograph records of events and tape recordings of interviews, meetings, and speeches.

6. *Remains or relics*. Fossils, skeletons, tools, weapons, clothing,

buildings, furniture, utensils, art objects, teaching materials, samples of examination question papers, samples of student work, and murals.

The remains or relics are more trustworthy than the records, reports or words of others, because a researcher can examine them personally. He can actually view, handle, measure, weigh, and describe an object or instrument of the past that has been preserved. However, he has to search for clues in reports made by men in the past to interpret how, when, or why the object or an instrument was used. These remains or relics are handed down from the past without the specific purpose of transmitting information from an unconscious testimony of events in the lives of people. They sometimes reveal the events and conditions better than official documents. The remains of an ancient burial place, for example, may impart more information about the food, tools, religious beliefs, customs and means of livelihood of a particular community.

Secondary Sources

Secondary sources are the accounts of an event provided by a person who did not directly observe the event, object, or condition. The person may have directly contacted an actual observer and talked with him or read an account by an observer. Since the testimony of the person is not that of an actual participant or observer, secondary sources are subject to an inherent danger of inaccuracy and distortion. For this reason, the researcher should rely as much as possible on primary sources and use the secondary sources only to bridge the gaps between the various pieces of primary data.

At times, however, it is not always possible to obtain primary data and in such situations the researcher may have to rely on secondary sources. These situations, according to Mouly (1963, p. 208), are frequent in education where only fragmentary reports concerning the processes of education are available. He is of the opinion that people in the past considered education so trivial that they did not bother recording anything about its nature or its organization; and consequently, it is relatively difficult to identify suitable primary data to permit the conduct of a good historical research in education. The personal documents as diaries and personal letters also leave wide gaps for the researcher to get the required continuity without resorting to secondary sources.

Secondary sources, if used carefully, serve many useful purposes. They may acquaint a researcher to major theoretical issues in his field and to the work that has been done in the area under study. They may suggest possible solutions of the problem and working hypotheses and may introduce the researcher to important primary sources.

A rigid classification of source materials is not always possible and practicable. Some type of data may be primary sources for some purposes and secondary sources for another. For example, a high school textbook in Indian History will be classified ordinarily a secondary source. But if one were making a study of the changing emphasis on national integration in high school history textbooks, the book would be a primary source of data.

In the location of source materials in historical research, the card catalog, periodical indexes, bibliographies, historical reviews, dissertations, and research journals provide helpful guides. It may be noted that historical studies involve more intensive bibliographical work and library usage than any other type of research, and hence, the researcher should be careful while assembling full bibliographical information in his note-taking system to facilitate proper documentation.

Criticism of Data

After the data have been identified, the researcher must learn to read them correctly as a basis for developing sound ideas of the past, which in turn may help in interpreating present trends and possibly in predicting future events. For this the researcher subjects his data to rigorous evaluation, which is known as criticism of the data. It involves the dual processes of establishing the authenticity of the source and of establishing the validity of its contents.

The process of establishing the authenticity of the data is termed as external criticism and that of establishing the validity of their content is termed as internal criticism.

External Criticism

External criticism, also called as lower criticism, checks the genuineness and authenticity of the source material. It helps to determine whether it is what it appears or claims to be and whether it reads true to be original so as to save the researcher from being the victim of a fraud. The purpose of external criticism according to Mouly (1963, p. 210), is, however, not so much 'negative' that is, the detection of fraud—as it is the 'establishment of historical truth'.

To determine the genuineness of the historical data, a researcher must possess a rich fund of historical and general knowledge. According to Van Dalen (1973, p. 168) he also needs "a good chronological sense, a versatile intellect, good common sense, an intelligent understanding of human behaviour, and good plenty of patience and persistence". The problem of establishing age or authorship of a document may involve such techniques as authentication of signatures, handwriting, script, and type; chemical analysis of paint, carbon dating of artifacts, ink, paper, cloth, stone, metals or wood. The researcher, therefore, must be familiar with chemistry, archaeology, cartography, art, literature, philology, anthropology, paleography, or various modern and ancient languages. If he does not have a knowledge of these fields, he may acquire special training in the fields that are most closely related to his historical problem or may seek the help of competent experts in the field.

Internal Criticism

After the authenticity of the historical data has been established, the researcher proceeds to internal criticism. It is also called as higher criticism and is concerned with the validity, credibility, or worth of the content of the document. Besides the textual criticism, it also involves such factors as competence, good faith, bias and general reputation of the author. Internal criticism is positive in nature when the researcher seeks to discover the literal and the real meaning of the text. It is negative when the researcher tries to seek every possible reason for disbelieving the statement made, questioning critically the competence, truthfulness or accuracy, and honesty of the author. Good, Barr and Scates (1941, p. 262) are of the opinion that both positive and negative criticisms are essential in historical research but the researcher should not go so far as to be cynical and hypercritical.

The competence and accuracy of an author is evaluated in relation to his status as a trainined eye witness, presence of emotional stress or pressure that might influence the observation, and the extent to which the conditions for observing were favourable. It is also evaluated in terms of the time period that has elapsed between the event and its recording by the author so as to ascertain whether the author was able to remember accurately the account of the event.

The author of a document may know the truth, but for certain reasons he may report the evidence only in part or in a distorted form. Distortion of the fact may result from author's motive, bias or prejudice. It may also result from his personal vanity or ambition, literary artifice, known ignorance about the subject, known weakness of telling lies or half truths, desire to flatter his superiors, desire to please the public, political or religious views, or vested interests. According to Good (1966, p. 173):

The test of truthfulness and honesty include evaluation of the author's characteristics and statements in relation to personal or vested interests, race, nation, party, region, sect, social level, economic group, profession, conventional formulas rather than true sentiments, vanity or boasting, attempt to please some individual or group exaggerations, and embellishments.

The validity of a historical fact contained in a document can sometimes be evaluated by comparing it with the statements of other authors. When there is disagreement among authors, the researcher must establish which one is correct. This he must do on the basis of overall credibility, reputation, independent authentication and general consistency with other known facts.

Good, Barr and Scates (1941, p. 259) have listed the following series of questions which a researcher may find useful in the processes of external and internal criticism of historical data:

1. Who was the author, not merely what was his name but what were his personality, character, position, and so forth?
2. What were his general qualifications as a reporter—alertness, character, bias?
3. What were his special qualifications and disqualifications as a reporter of the matters here treated?
    (*a*) How was he interested in the events related?
    (*b*) How was he situated for observation of the events?
    (*c*) Had he the necessary general and technical knowledge for learning and reporting the events?
4. How soon after the events was the document written?
5. How was the document written, from memory, after consultation with others, after checking the facts, or by combining earlier trial drafts?
6. How is the document related to other document?
    (*a*) Is it an original source; wholly or in part?
    (*b*) If the latter, what parts are original; what borrowed; whence? How credible are the borrowed materials?
    (*c*) How and how accurately is the borrowing done?
    (*d*) How is the borrowed material changed; how used?

Invariably, the researcher gets answers for all these questions and, therefore, he has to depend upon partially, on evidences he can no longer verify. At times, he will have to rely on the inferences based upon logical deductions in order to bridge the gaps in the information. In such cases, Moulay (1963, p. 214) suggests that a researcher should predicate his remarks with the phrase "according to..." in order to safegaurd his reputation and status as a scientist.

Interpretation of Data

After the data have been collected and criticized, the researcher turns himself to the task of interpretation of these data in the light of his problem. Although the interpretation of data is treated earlier in the Chapter 6 , a brief discussion may be made at this stage, with special application to the historical data. Because of the unique nature of the historical data, the task of interpretation becomes complicated and acquires special significance. It requires greatest ingenuity and imagination on the part of the researcher.

The researcher in the historical type of investigation must be very cautious while dealing with 'cause and effect' relationships. Here his position is entirely different from a researcher of physical sciences who deals with very simple isolated laboratory phenomena. Historical causes are invariably complex and the historical researcher must accept the fact that he is not dealing with clear-cut cases of causes and effect. According to Mouly (1963, p. 214).

> Causation is a troublesome concept in science; it is doubly so in historical research where 'causes' are in the nature of antecedents, or precipitating factors, rather than 'causes' in restricted scientific sense.

Since history is actually a record of the chain of related events, it becomes very difficult for a researcher to interpret that one event in the chain was caused by the previous event in the chain. Furthermore, many conditions and circumstances interact and become responsible for a particular event. Therefore, it makes the task of the researcher difficult, to assess accurately the influence of a particular event and to identify clear-cut cause-effect relationships.

The historical researcher must also be very cautious in his use of analogy in the interpretation of data. While drawing comparisons between one historical event and any number of others, he must carefully make use of similarities as well as of differences.

The ultimate goal in the historical type of research is not only to establish facts but also to determine trends which the data may suggest and to draw inferences from the data. The researcher must show an understanding of the sequence of events and must draw a vertical relationship of proceeding facts with succeeding ones along the time line. His goal should be one of synthesis and interpretation rather than mere summation.

Writing of the Research Report

After the data have been interpreted, the researcher has to write a well-organized report of his study. Since a detailed discussion of the processes involved in report writing is presented in Chapter 9, it is sufficient to state here that the report of the historical research includes a statement of the problem, a review of the literature, the statement of the objectives and hypotheses, survey and sources of data, and methods of its collection, organization of heads of classification and ordering of data; the criticism; analysis and interpretation of data, the conclusions reached and a bibliography.

The writing of historical research report needs the highest level of scholarship on the part of the researcher. This is a matter of expositional strategy which calls for creativity in addition to the qualities of imagination and ingenuity. The researcher must be elegant and objective in his style of writing the research report. However, because of the discontinuous and incomplete nature of historical data on which valid generalizations can be established it is generally accepted that in the writing of historical research report, the researcher has to be permitted a little more freedom on the subjective interpretation of data. At the same time, it must be ensured that the presentation is sufficiently systematic and does not hide or distort the truth.

At times, the gaps in the historical data are such that an attempt at interpretation becomes futile and unsafe. In such situations the best course for the researcher is to indicate the existence of the gap and stop there, or to suggest a number of alternative explanations.

The historical research report must be presented in the logical, chronological and topical order. Good, Barr and Scates (1941, p. 265) also recommend that an appropriate combination of the chronological and topical organization of historical data, involving consideration of such influences or forces as political institutions law, economics, geography, social conditions, war, national culture, art, literature, religion, great leaders, natural sources, etc. seems best.

Reports of historical research should neither be dull and unattractive nor too flowery and ornamental. They must follow precision, continuity, clarity and dignity in their style to give a sense of design and completeness.

POST TEST

1. Describe briefly the importance of research method in the research process.
2. Name the three basic categories of research methods.
3. Define the term history.
4. Describe the nature of historical research.
5. Differentiate between history and historical research.
6. Describe the value of historical research in education.
7. List and describe the types of historical research.

8. State a topic of historical research in education and discuss the various steps which you would take in investigating it.
9. Explain, by giving examples, the meanings of primary and secondary sources of data, and external and internal critisms of historical data.
10. How is historical evidence validated in historical research?

11. The Descriptive Method

This chapter has four sections. The first section deals with the nature of descriptive research. The second section describes briefly the value of descriptive research in education. The third section is divided into four sub-sections that describe in detail the four categories of descriptive research: survey studies, interrelationship studies, developmental studies, and content analysis. The first sub-section deals with the nature of survey studies and describes in detail their various types: school surveys, job analysis, documentary surveys, public opinion surveys, and social surveys. The second sub-section deals with the nature of interrelationship studies and describes in detail their various types: case study, causal-comparitive studies, correlation and prediction studies, and cross-cultural and comparative studies. The third subsection deals with the nature of developmental/genetic studies and describes in detail their various types: growth studies (longitudinal and cross-sectional studies), follow-up studies and trend studies. The fourth sub section deals in detail with the nature, utility and procedure of content analysis. The fourth and the final section of this chapter describes briefly the steps which the researcher may adopt in conducting the descriptive type of research.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Describe the nature of descriptive research.
2. Describe the value of descriptive research in education.
3. List and describe the various types of descriptive research.
4. Describe the nature of survey studies.
5. List and describe the various types of survey studies.
6. Describe the nature and purpose of school surveys.
7. List and describe the various types of school surveys.
8. Describe the nature and various types of survey testing.
9. Describe the uses of achievement testing, intelligence testing and personality testing in educational research.
10. Describe the nature and purpose of school appraisal.
11. Describe the nature and purpose of status studies.
12. Describe the nature and purpose of financial studies.
13. Describe the nature and purpose of curriculum studies.
14. Describe the nature and purpose of building surveys.
15. List and describe the various ways in which the school surveys can be organized.
16. Describe the nature and purpose of job analysis.

17. Describe the nature and purpose of documentary analysis.
18. Describe the nature and purpose of public opinion surveys.
19. Describe the nature and purpose of social surveys.
20. Describe the nature of interrelationship studies.
21. List and describe the various types of interrelationship studies.
22. Describe the nature of case study.
23. List and describe the steps of the case study method.
24. List and describe the advantages of case studies.
25. List and describe the limitations of case studies.
26. Describe the nature of causal comparative studies.
27. List and describe the uses of causal comparative studies.
28. List and describe the limitations of causal comparative studies.
29. Describe the nature of correlation studies.
30. List and describe the various types of correlation analyses.
31. Describe the method of interpreting the results of a correlation study.
32. Describe the nature and purpose of prediction studies.
33. Describe the nature of cross-cultural and comparative studies.
34. Describe the uses of cross-cultural and comparative studies.
35. List and describe the limitations of cross-cultural and comparative studies-
36. Describe the nature of development, genetic studies.
37. List and describe the various types of developmental studies.
38. Describe the nature of growth studies.
39. List and describe the two types of growth studies.
40. Describe the nature and purpose of longitudinal studies.
41. List and describe the limitations of longitudinal studies.
42. Describe the nature and purpose of cross sectional studies.
43. List and describe the limitations of cross sectional studies.
44. Describe the nature and purpose of follow-up studies.
45. Describe the nature and purpose of trend studies.
46. Describe the nature of content analysis.
47. Describe the utility of content analysis in education.
48. Describe the procedure of content analysis.
49. List and describe the steps which the researcher may adopt in conducting the descriptive type of research.

It has been pointed out earlier in Chapter 10 that choice of the method of research is determined by the nature of the problem. Historical methods can tell us much about what existed in the past by determining, evaluating, and understanding past events; descriptive methods can tell us about what exists at present by determining the nature and degree of existing conditions. Because of the method's apparent ease and directness, descriptive method has undoubtedly been the most popular and most widely used research method in education.

NATURE OF DESCRIPTIVE RESEARCH

Descriptive research studies are designed to obtain pertinent and precise information concerning the current status of phenomena and, whenever possible, to draw valid general conclusions from the facts discovered. They are restricted not only to fact finding but may often result in the formulation of important principles of knowledge and solution of significant problems concerning local, state, national and international issues. Descriptive studies are more than just a collection of data; they involve measurement, classification, analysis, comparison, and interpretation. They collect and provide three types of information : (1) of what exists with respect to variables or conditions in a situation; (2) of what we want by identifying standards or norms with which to compare the present conditions or what experts consider to be desirable, and (3) of how to achieve goals by exploring possible ways and means on the basis of the experience of others or the opinions of experts.

The activities of descriptive studies researchers are not different from those of the other researchers. As in any study they (1) identify and define their problem; (2) state their objectives and hypotheses; (3) list the assumptions upon which their hypotheses, and procedures are based; (4) choose appropriate subjects and source materials; (5) select or construct tools for collecting data; (6) specify categories of data that are relevant for the purpose of study, and capable of bringing out significant similarities, differences, or relationships; (7) describe, analyze, and interpret their data in clear and precise terms; and (8) draw significant and meaningful conclusions. At times, descriptive researcher is not directed towards hypothesis testing. For example, a District Education Officer wishes to know the number of the students placed in first, second and third divisions in matriculation examination in a particular year of a district school-wise. He is merely interested in the number, wants to describe a situation that prevails and attempts no explanation of underlying reasons responsible for turning out a large number of first divisioners by some school/schools. This study does not propose or test hypotheses and does not deal with many variables underlying the achievement or lack of achievement of students. It may merely provide information to assist in decision making.

Descriptive studies investigate phenomena in their natural setting. Their purpose is both immediate and long range. They constitute a primitive type of research and do not aspire to develop an organized body of scientific laws. Such studies, however, provide information useful to the solution of local problems and at times provide data to form the basis of research of a more fundamental nature.

Descriptive research differs from other types of research in purpose and scope. A clear-cut distinction can be drawn between descriptive studies and historical studies on the basis of time. The latter deals with the past and the former with the present. The limitations of descriptive investigations, however, are very similar to those of the historical enquiry in that cause and effect relationships are difficult to establish, and the time at which study is conducted is a critical factor in the interpretation of the data. The method of descriptive research, in contrast to an experiment, is relatively less scientifically sophisticated. Here the researcher does not manipulate the variables or arrange for events to happen. In fact, the events that are observed and described by him would have happened even though there had been no observation. Descriptive studies involve events that have already taken place and are related to a present condition.

Descriptive studies vary greatly in complexity. At one extreme, they constitute nothing more than frequency count of events to the study of local problems without any significant research purpose. At the other extreme, they attempt to ascertain significant interrelationships among phenomena.

VALUE OF DESCRIPTIVE RESEARCH IN EDUCATION

The descriptive research method has undoubtedly been the most popular and the most widely used research method in education. It helps to explain educational phenomena in terms of the conditions or relationships that exist, opinions that are held by the students, teachers, parents and experts, processes that are going on, effects that are evident, or trends that are developing. Because of the apparent ease and directness of this method, a researcher can gather information in terms of individual's opinion about some issue, by a simple questionnaire. At times, descriptive survey is the only means through which opinions, attitudes, suggestions for improvement of educational practices and instruction, and other data can be obtained.

The descriptive investigations are of immense value in solving problems about children, school organization, supervision and administration, curriculum, teaching methods and evaluation. There are a number of questions that arise concerning these aspects of education. For example, one may want to know how many of the teachers in a district possess a Bachelors degree in Education. How do these figures compare with the tendency throughout the state? How many minutes per week are normally devoted to the teaching of English spelling? What proportion of total state budget is it reasonable to set aside for adult education? What kind of a curriculum do people really want their children to have at the secondary school stage? At what age and grade level do pupils leave school? What happens to children after they leave school? What higher institutions or vocations do they enter? and so on. Such information is useful to teachers and administrators and in understanding the existing educational problems and also in suggesting ways of meeting them. The head of a school may wish to know how other school systems are being run, so that he can compare his practices with theirs. This way he will be able to know what procedures and standards are superior to those of other schools. The teachers will also study the conditions existing in their classrooms and that of other teachers.

The problems in education directly involve people and the situations precipitating these problems are constantly in a state of change. To keep abreast of changes, descriptive studies conducted at different intervals with representative groups of people will be immensely helpful.

The descriptive type of research is useful in the development of data gathering instruments and tools like checklists, schedules, score questionnaires and rating scales. It also provides the background ideas and data from which many more refined or controlled studies of casual relations are made.

TYPES OF DESCRIPTIVE RESEARCH

Descriptive studies have been classified variously by various writers. Some have classified them on the basis of the purposes they achieve; some on the basis of the geographical areas they cover; and some on the basis of the techniques they employ. These classifications mostly range from the survey, which describes the status quo of educational variables, to the correlational study, which investigates the relationships between variables. For the sake of convenience descriptive studies may be classified in the following three categories:

1. Survey studies,
2. Interrelationship studies,
3. Developmental studies

Although this classification is arbitrary, yet it seems to have merit from an operational, as well as from an organizational point of view and will be used as the basis of the present discussion. Some investigations fall exclusively within one of these categories, but others hav characteristics of more than one.

Survey Studies

Survey studies are conducted to collect detailed descriptions of existing phenomena with the intent of employing data to justify current conditions and practices or to make more intelligent plans for improving them. Their objective is not only to analyze, interpret, and report the status of an institution, group, or area in order to guide practice in the immediate future, but also to determine the adequacy of status by comparing it with established standards. Some surveys are confined to gather all three types of information: (1) data concerning existing status, (2) comparison of existing status with the established status and standards, and (3) means of improving the existing status;,while others are limited to one or two of these types.

Survey studies may take different forms depending upon the scope, nature and purpose of the problem under investigation. They may be broad or narrow in scope. Some surveys encompass several countries, states or regions; or may be limited to one country, region, state, district, tehsil, city, school system, or some other unit. Survey data may be collected from every unit of a population or from a representative sample. The information gathered may be concerning a large number of related factors or may be confined to a few selected items.

Survey studies describe and specify the properties of educational phenomena. They include: (1) School surveys, (2) Job analysis, (3) content analysis; (4) Public opinion surveys, and (5) Social surveys.

1. *School surveys.* The school survey is not new. The first formal survey of the schools of Boise, Idaho (USA) was conducted in 1910. This survey was largely of an inspectional nature and it gradually became the pattern for other school surveys in the United States. But these surveys generated apprehension as well as opposition on the part of local teachers because of their inspectional nature. Furthermore, it was felt that such surveys generally lacked continuity due to non-implementation of the recommendations, and, therefore, they were of limited overall value. Consequently, some improvements were suggested and emphasis now is on comprehensive surveys designed to evaluate the school as a functional unit.

Nature and purpose of school surveys. The school survey generally is a comprehensive study of existing conditions. Its main purpose is to determine the overall effectiveness of the school programme and suggest improvement where necessary. In terms of purposes, Cooper (1946) and Reller (1942) have categorized school surveys into three types: (*a*) the investigative, evaluative or status survey which serves primarily to evaluate existing conditions; (*b*) the deliberative, developmental, or planning survey, which is intended primarily to make proposals for development and improvement, with a minimum criticism of present circumstances; (*c*) the implementive survey, which not only makes suggestions for development but also attempts to create conditions in the conduct of the survey which will enhance the prospects of actually achieving survey recommendations.

According to Good (1966, p. 207), the comprehensive school survey usually covers the following aspects of the school system:

1. Aims, outcomes, pupil achievement, curriculum, method, and instructional aids.
2. Administrative problems and procedures of the schools.
3. Financial policies and procedures.
4. Operation and maintenance of the physical plant.
5. Pupil transportation.
6. Staff and personnel.
7. School plant and related factors.

School surveys may be conducted at the local, state, regional or national level. They may be undertaken at various levels of instruction—elementary, secondary, or higher. Some of them may be concerned with the problem of school building only. Others may study community and school setting, available financial resources, pupil transportation system, or estimate future school enrolment.

Acquiring information about the characteristics of educational personnel and nature of pupils may be the purpose of many school surveys. Such surveys may gather information about teachers, supervisors, and administrators. They may study the behaviour of teachers and other instructional personnel in the classroom, the school, and the community, with the objective of assessing or improving teaching effectiveness. Determining the physical-fitness of teachers, their attitudes, socio-economic conditions, responsibilities, status and salaries may be the objectives, of some surveys. Information about the behaviour patterns of pupils in classrooms, with peers, at home, and in the community may also be sought through the school surveys. Some surveys are concerned with the study of pupils' knowledge, skills, academic achievement, intelligence, aptitudes, work or study habits, health practices, extra-curricular activities, and recreactional aspects. School surveys may also explore the number of drop outs, the number and type of handicapped or other exceptional students, or the number and nature of delinquency incidents.

School surveys are sometimes conducted to study the educational programmes, processes, and outcomes. They may seek information about the equipment and supplies, such as the number and kinds of library books or the amount and types of laboratory, athletic or audio-visual equipment. Some surveys may be concerned with the curriculum, textbooks and other instructional materials. Other surveys may seek information about aspects of social structure in the classroom, home, or community that may influence learning.

The scope of school surveys is large and varied. A single comprehensive school survey may be comprised of various parts or constituent surveys. These include: (1) survey testing; (2) school appraisal; (3) status studies; (4) financial studies; (5) curriculum studies; and (6) building surveys.

Survey testing. Tests serve a variety of functions. They are intended not only to measure maximum capability, such as achievement, aptitude and intelligence but also to describe typical behaviour such as personality traits, interests and attitudes. For the purposes of survey testing, we shall be mainly concerned with achievement, intelligence and personality testings.

Achievement testing. Survey testing of educational achievements has become an important aspect of school surveys. Achievement tests of the objective type are being, and have been constructed by various agencies. They are based upon the current courses of study prescribed by school boards for the schools within the area or system under investigation and may cover any or all subjects and grades. Achievement testing surveys may serve the following purposes:

(*a*) They may be used to compare the attainment of present pupils with that of the previous years' and may in the case of unsatisfactory performance lead to an investigation of the causes and remedies.

(*b*) They enable the administrator to compare the performance of different schools.

(*c*) They furnish national or local norms which can provide a basis for assessing certain curricular strengths and weaknesses.

(*d*) They help to determine whether an innovative programme is better or poorer than the conventional one in facilitating the attainment of specific curricular objectives.

(*e*) Teachers and heads of the schools may use the results of city or state-wide survey testing for a critical analysis of their own classes and schools.

(*f*) They enable to identify particular groups of students who are weak so that appropriate remedial measures could be introduced to bring them up.

(*g*) They provide information regarding the objectives of education which are fulfilled satisfactorily and which are not, so that changes in the methods of teaching could be introduced to improve the quality of education and thereby ensure realisation of the desired objectives.

(*h*) They help in identifying particular groups of students in a state or the nation who are doing better so that studies could be conducted to investigate the characteristics of the educational system being followed there and make the results known, for the benefit of others.

(*i*) They help in locating geographical regions or school systems which are weak so that they could be given more help.

Achievement testing surveys are conducted not just to collect some figures but to provide necessary information for educational administrators and teachers for taking decisions and actions to improve instructional process. A number of such surveys covering large areas have been conducted in the USA. In India, the first step towards such a comprehensive study was undertaken by National Council of Educational Research and Training New Delhi (Kulkarni *et al.*, 1968). The survey was confined to mathematics achievement and conducted at the three levels of education, primary, middle, and high school levels respectively, in all the states of India (except Bihar and Tamil Nadu). Many other organizations like university teaching departments of education, state institutes of education, etc. are constructing achievement tests for the purpose of achievement testing surveys.

Intelligence testing. The intelligence tests are useful tools for educational researchers. In large group studies of the survey type, the researchers invariably make use of group tests of intelligence—verbal or non-verbal, or both. The purposes for which the intelligence-testing surveys may be used are the following:

(*a*) For identifying gifted children.

(*b*) For studying mental growth.

(*c*) For making decisions for individual students regarding their sucess in college, university or some other professional course.

(*d*) For measuring general readiness for learning at different levels of schooling.

(*e*) For dividing large groups of students into relatively homogeneous sections.

(*f*) For estimating more accurately the degree of mental retardation among children.

(*g*) For constructing or adopting intelligence tests.

(*h*) For studying the socially and educationally maladjusted children.

(*i*) For indicating the extent of differences of IQ among the children of same chronological age.

Personality testing. Personality in a broad sense is a general term covering the entire constellation of personal characteristics. It can be defined as including attitudes, interests, and various other characteristics such as traits, adjustment and needs. Questionnaires, interviews, observations, checklists, rating scales, attitudes scales, personality and interest inventories are frequently used for survey testing in the field of personality.

School appraisal. Appraisal of different aspects of a school or school system is an essential part of school surveys. It is a form of evaluation or scaling according to subjective values. School appraisal is concerned with administrative provisions and practices of an educational institution and the educational attainments of its pupils. It aims at an evaluation of all the conditions and outcomes in a school system. Besides checklists, rating scales, or score cards, the researcher frequently employs other types of evidence such as academic achievement and subsequent success of the pupils. In the USA, a number of schedules and checklists for evaluating schools in such areas as English, physical education, agriculture, industrial arts, health and home economics, social studies, pupil activity projects, guidance, school staff and administration, library service, etc. have been prepared and used.

Status studies. Status studies are undertaken for determining the personal and professional status of teachers and other school officials. They may be a part of a school survey, or form a subject for independent study. Studies to investigate the problems of selection and placement of teachers, their work load, their status in relation to the service conditions, health, law, supply and demand, etc. all provide useful subjects for research. In addition to questionnaires and schedules, official records are also utilized in status studies.

Financial studies. The problems relating to the financial position of a school in terms of the sources of finance, the items of expenditure, the deficit, expenditure per student and salaries of teachers form important aspects of school studies. Official records are the main sources for providing data. Questionnaires are also employed by the researchers.

Curriculum studies. The analysis of the existing curriculum in a school system and its comparison with the one existing elsewhere may be a part of a school survey. It can also be taken as an independent study. The curriculum analysis usually includes an analysis of the principles on which the curriculum is based, the needs of the pupils and the community it fulfils, and its limitations. Questionnaires, schedules and checklists are generally employed in these studies.

Building surveys. Surveys of school buildings are undertaken to gather information about their sites and location, the accommodation they provide and the arrangements they have for various classes, subjects, activities or staff. They also provide information about the condition as regards their maintenance and cleanliness, and the improvements or expansions they stand in need of. The building survey of a school system, according to Harris (1960, p. 1021), consists essentially of three steps: (*a*) the determination of building facilities needed; (*b*) an assessment of the existing and possible sources to accommodate the needs; (*c*) specific recommendations as to new structures, their general sizes, functions, and locations, and renovations of existing structures. Building studies may form a part of comprehensive school surveys or they may be independent in nature. Questionnaire, checklist, score card or observation are usually employed as tools for collecting data in such studies.

Organization of school surveys. The school survey, according to Mouly (1963, p. 311) can be conducted in one of the given three ways: (1) by outside consultants (the consultant survey); (2) by the personnel of the local school system (the self survey); or (3) by a comprehensive group consisting of local teachers, interested members of the community, and resource persons, headed by an expert acting as consultant and survey leader (the comprehensive survey).

The consultant survey. The consultant survey is conducted by the experts or specialists outside the school. It is useful when the teachers of the school are unable to cope with a problem or when the problems are so comprehensive that they create an overburden on the teachers. The consultant conducting the survey has an advantage in that he can see the school in the light of his previous experience in similar schools. He is more objective and alert to the problems that may have become 'blind spots' to those in the school system. The consultant frequently can command more prestige and respect by virtue of being an outsider. He usually has highly specialized training in research and, hence, undoubtedly he is a capable person to conduct a survey meaningfully and profitably.

The consultant survey is, however, subject to certain limitations. Since a survey generally is of comprehensive nature, a large number of weaknesses in the school system will probably be found. Anticipating this, the personnel in the system are reluctant to cooperate in the discovery of their weakness. Furthermore, the teachers and other personnel have almost no active participation in the survey, they are likely to be only mildly enthusiastic in implementing its recommendations.

The self survey. The self survey is usually conducted by the teachers of a local school system. This method is subject to certain serious limitations. The teachers generally lack the insight into the true nature of their problems and the necessary competence to deal with them. Furthermore, the teachers are so close to their own problems that at times they consider them so trivial in nature that they are no longer considered problems, or so ingrained that they are considered beyond solutions. The conduct of self surveys add more responsibilities and extra work to the teachers who are already over burdened with their routine teaching work and, hence, they are more likely to resent the extra work than they are to look for the solution of the problems.

The comprehensive survey. In view of the limitations of the consultant survey and the self survey, the comprehensive survey is regarded as the best approach to school evaluation. In this method, the school supplements its own personnel by enlisting the cooperation of teachers from neighbouring schools, interested leaders of the community, and specialists from the universities and other institutions.

The comprehensive survey method has several psychological, academic and administrative advantages. The participation of teachers in the survey helps in identifying the problems of the school as well as in the implementation of the recommendations. Furthermore, the active involvement of selected and interested community leaders promotes community good will by acquainting its members with the philosophy of the school, its problems, and the limitations under which it functions. The specialists from the neighbouring university with their highly specialized training in research and wide experience frequently bring a fresh approach to the problems of school.

2. *Job analysis.* The method of job analysis is generally used in business and industry. In education, it is employed to gather information about the general duties and responsibilities of the teaching, nonteaching and administrative personnel, the specific duties that they perform, their working conditions, the nature and type of their facilities, and their status and relationship in the administrative organization. The data about the education, specialized training, experience, skills, habits, health standards, and behavioural traits of these personnel are also collected. These data help researchers to get knowledge about the existing practices and conditions of employment, and the competencies and behavioural traits that the personnel possess or should possess to carry out their work effectively and efficiently. Van Dalen (1973, p. 200) has pointed out that an analytical knowledge of job components helps administrators and educationists to:

1. detect weaknesses, duplications, or inefficiency in the present work procedures,
2. establish uniform classifications of similar work,
3. determine wage or salary schedules for jobs entailing various levels or skills or responsibility,
4. identify the competencies to seek when employing personnel,
5. assign workers to jobs in a manner that will achieve the best utilization of the available manpower,
6. set up training programmes, and prepare instructional materials for prospective or inservice employees.
7. establish requirements for promotion,
8. make decisions concerning the transfer or retraining of personnel, and
9. develop theoretical framework for studying administrative functions and structures.

The researchers make use of various tools and techniques to collect objective and reliable data while undertaking job analysis. The may interview workers to seek information concerning the nature of their duties, training, or professional activities. They may make use of a checklist compiling the broad functions involved in administrative, supervisory, or teaching positions.

3. *Content analysis*. Content analysis is closely related to historical research. Both methods of investigation require that researchers study existing records, but historical research is primarily concerned with the more distant past and descriptive research with the present. Since content analysis is concerned with the classification quantification, and comparison of the content of document or communication, it is sometimes referred to as 'documentary', 'activity' or informational' analysis. The communication may be in the form of official records (census, birth, accident, crime, school, institutional, and personal records) or any written, printed, verbal, or pictorial form.

Content analysis may be of various types. Researchers may analyse textbooks, syllabi, courses of study, lesson plans of pupil teachers, and prospectus of various universities and colleges. They may analyze the contents of reference works, newspapers, periodicals and journals. Sometimes study of personal documents, such as diaries, letters and notes, provide researchers valuable information. Researchers may also examine and analyze judicial decisions and state laws; administrative records, forms and reports; budgets and financial records; and cumulative records.

Utility of content analysis in education. Every classroom teacher uses content analysis informally to determine the level of vocabulary in a particular textbook, to describe the differentiating aspects of different writing styles, to know the possibility of bias in the historical writing to choose textbooks for the use of his students, to use cumulative records as a basis for decisions about students, to assign marks to essay-type questions, and the like. The decisions which are based on informal content analyses lack reliability and validity because of high subjectivity and the absence of any systematic methodology. To obtain reliable and valid results formal content analyses can be undertakin in the following areas:

1. *Curriculum.* Content analysis can be used in developing and modifying school curriculum at various stages. Theachers too often experience that the school curriculum is not specifically defined and it does not satisfy the set goals or objectives. Curriculum research can be conducted through content analysis of textbooks, pamphlets, reference books and other written material.

2. *Evaluation.* Another important use of content analysis is in the field of evaluation. Evaluation helps the teacher to know whether the objectives have been met by the curriculum. For this he has to develop tests or make use of readymade standardized tests. If the teacher has to develop his own tests, he has to spend a lot of money and effort. He will be at considerable advantage if he is able to locate suitable test which has been already developed on the basis of the curriculum he is following in the school. To determine the extent to which a particular readymade standardized test is appropriate, it is essential that the test must be content analysed by some means. The test items can be compared to the objectives and desired outcomes of the curriculum and the magnitude of agreement can be established.

3. *Comprehension level of students.* It is essential for the teacher to ascertain that materials used in the textboooks are at a comprehension level appropriate for the students. Content analysis can also be used to determine the comprehension level of students using the sophisticated readability techniques or formulae developed by Dale and Chell (1948) and Flesh (1951). Such techniques or formulae are useful for arriving at a number which describes readability of comprehension level of some written material.

Procedure of content analysis. The procedure involved in conducting a content analysis is more or less different from other forms of descriptive research. Hayman (1968, pp. 80-82) has specifically summarized the procedure as under:

1. *Defining the unit of analysis.* The first step in preparing for a content analysis is to define the unit of analysis. The researcher

must decide in advance whether the unit is confined to single words, to phrases, to complete sentences, to paragraphs, or to even larger amounts of materials such as articles or complete books. Either of these can be considered as an entity whose specified characteristics can be determined and analysed.

Hayman (1968, p.80) is of the opinion that a 'unit of meaning' should be used. He suggests that the unit should be comprehensive enough to provide meaning through some context at least, but small enough not to allow subjectivity in its use.

Once ,the unit is defined, the researcher will provide for classifying units in the material to be analyzed according to frequency, direction, intensity, and contingency.

2. *Frequency*. The researcher merely counts the number of units which fall into each of his categories for determining the frequency.

3. *Direction and intensity*. In certain situations, it is worthwhile to further classify the unit according to direction and intensity.

Direction refers to whether the reference was favourable, unfavourable, or neutral. It might be pleasant-unpleasant, interesting-uninteresting, threatening-non-threatening.

Intensity indicates the emotional impact of the unit analysed. Is it great or small, and in what direction?

Judging direction and indensity is more subjective than merely counting for frequency.

4. *Contingency*. The contingency analysis aims at considering the context within which the unit is found. A researcher should consider the favourableness or unfavourableness of a single unit in the light of the remainder of the communication so that its real meaning might not be lost.

5. *Sampling*. One of the major and practical problems in content analysis is sampling. The unit which a researcher analyses must be representative of the total meterial with which he is concerned, so that the results can be generalized. The researcher should not be over ambitious in selecting the unit. Very seldom will he be able to analyse the total unit in which he is interested. If he is concerned only with a single book, he can at best analyse only part of the material in it.

6. *The recording instrument*. A form is usually used in content analysis for classifying and recording of the units. The researcher should use the form carefully in order to get reliable and vailed data. He should clearly indicate the categories and make provision for marking the category into which each unit falls. Their rating scales

4. *Public opinion surveys.* In order to make some important and crucial decisions, industrial, political, educational and other leaders seek knowledge of the public's opinions, attitudes, and preferences. Educationists, for example, conduct public opinion surveys to find out how people feel about school issues. Business firms make public opinion surveys to find what type of products, packaging, or advertising appeals to purchasers. Politicians conduct opinion surveys to ascertain how people will vote or what programmes they favour.

In public opinion surveys, the researchers usually make use of questionnaires, schedules or interviews to gathers data from the selected group or groups following appropriate sampling procedures.

Public opinion surveys have several limitations. At times, the data are collected from a readily available group of people which may not always reflect the opinion of the total population. The situations and the environment in which the study is conducted may also affect the reliability of the data. Furthermore, it is difficult to measure the intensity or depth of opinion in such surveys.

5. *Social surveys.* Social surveys are also called community surveys. These surveys, while essentially the same from the stand point of the research procedures involved, differs from the school surveys in that they are more general and comprehensive. Since school is one of the many agencies whose function is vital to community welfare, social surveys may include data concerning schools, and conversely school surveys may analyze many aspects of the society or community. The social surveys are of great benefit to the school in clarifying the social setting in which they exist and function and the expectations of the community with respect to the education of its citizens.

Social surveys are generally undertaken to study health services, employment conditions, causes of juvenile delinquency, housing problems, or caste discriminations. Such surveys can be conducted by the state or local government, by local community leaders independently of the government, or by a group of experts financed by the government or some other agency.

Researchers who conduct social surveys employ research methods from various fields and gather information concerning many factors that contribute to the character of the community life. They make use of research tools like questionnaires, schedules, interviews, rating scales, and direct observation as well as statistical, ecological and other techniques to gather data from government officials, social agencies, community leaders, schools and other educational institutions, students and teachers, and various documents.

Interrelationship Studies

Some researchers in the field of education do not merely gather facts to obtain an accurate description of existing phenomena; they attempt to trace relationships between facts that will provide deeper insight into the phenomena. The studies that endeavour to discover relationship between various facts of the existing phenomena are called as interrelationships studies. These include: (1) case study; (2) causal-comparative studies; (3) correlation and prediction studies; (4) cross cultural and comparative studies.

1. *Case study.* The typical case study is an intensive investigation of a social unit. The social unit may be an individual, a family, a school, a group of delinquents, dropouts, or any teenage gang.

In a case study, the researcher attempts to examine an individual or unit in depth. He gathers pertinent data about the present status, past experiences, and environmental forces that contribute to the behaviour of the individual or social unit, and how these factors relate to one another. The analysis of the factors and their interrelationships, help the researcher to construct a comprehensive and integrated picture of the unit.

Nature of case studies. The case study method was originally used in medicine to examine the patient's previous development, his health and physical state from the beginning and many other factors in the past, besides making a careful study of the patient's present condition and symptoms. Freud used case study method to assist his subjects in solving their personality problems. The published detailed accounts of his interviews with patients and his interpretations of their thoughts, dreams, and actions provide excellent examples of case studies.

Guidance counsellors and social workers conduct case studies for diagnosing a particular condition or problem and recommending therapeutic measures. They gather data from a particular individual and confine their interest to the individual as a unique personality. Researchers, on the other hand, are interested in individuals as representative types. They gather data about a carefully selected sample of individuals so as to derive valid generalizations about the population that the sample represents.

A case study is not different to a survey, but instead of collecting information about few factors from a large number of units the researcher makes a depth and intensive study of a limited number of representative cases. It is narrower in scope but more exhaustive and more informative in nature than a survey. To provide more qualitative data, the case study is often used to supplement the survey method.

Though case studies in general are classified as descriptive research types, they have sometimes been conducted for purpose of hypothesis testing and taken the form of experimental research. Many case studies, for example, were conducted to investigate the effects of operant conditioning on human behaviour. In a typical study, as reported by Ary *et al*. (1972, p. 288), the researcher identifies a specific behaviour in his subject and systematically records the frequency of this behaviour. Then he introduces an operant conditioning treatment and records the frequency of the specified behaviour during treatment. When a change is observed in the behaviour of the subject as a result of operant conditioning, the researcher begins reversal conditioning; that is, he uses operant conditioning to change the behaviour back to what it was before the original conditioning was instituted.

The steps of the case study. The following steps are involved in the conduct of the case study:

1. The first step is to determine the present status of the individual or the social unit under investigation through direct observation or measurement. Here the researcher goes far beyond casual observation or superficial description. In addition to a physical examination of the subject, a psychological evaluation designed to determine the general ability level and the emotional maturity of the case is necessary. There are numerous standardized tools that are useful to the researcher in this assessment process. For example, to make a case study of a delinquent child, the first thing the researcher has to do is to survey the present status of the child by making an assessment of his physique, cognitive and non-cognitive factors through direct observation and administering tests of intelligence, aptitudes and personality.

2. The next step is to determine the most probable antecedents of the case and to formulate a fruitful hypothesis or a set of hypotheses through the knowledge of similar cases. The researcher, for example, can formulate a hypothesis that the occurrence of delinquent behaviour in a child is due to inadequate home environment, poor teaching in the school, low mental ability or any other factor.

3. The third step is verification of the hypothesis. The case is then checked for the presence or absence of the antecedents supposed to apply to the situation under investigation. Here the researcher makes use of the knowledge of the present status and the history of the case. He should not overemphasize observational methods and neglect other methods. Van Dalen (1973, p. 209) suggests that a multimethod approach may serve as a more valid test of a hypothesis. The re-

searcher may ask the case to recall past experiences or to express present wishes in interviews or questionnaires. Personal documents, such as diaries and letters, and various physiological, psychological or sociological measurements may provide valuable information. Data may be obtained from teachers, friends, parents, brothers, sisters and other family members.

4. After verification of the hypothesis, the next step is directed towards further validation of the diagnosis. Some remedial measures in the light of the causes found are suggested.

5. The last step of the case study is the follow up of the case. The case under study is re-examined to **ascertain** whether any changes have been produced by the treatments introduced. If the change is positive and significant, the diagnosis is taken to be correct.

Advantages of case studies. Case study method has the following advantages:

1. The case study attempts to understand an individual or a unit in depth. It tries to understand the whole 'case' in the totality of his environment. Not only the present status of an individual, but his past can be thoroughly probed.

2. The case study often provides an opportunity for a researcher to develop insight into basic aspects of human behaviour.

3. The case study helps the researcher to observe events both within and outside the educational setting in their totality. He notes events in terms of their location in space and time and interprets results in the light of the relationships and interdependence of events.

4. A case study may provide insights that will help a researcher to formulate fruitful hypothesis or a set of hypotheses. Freud, for example, when attempted to probe deeply into the dynamics of his patients problems, reasoned that the relationships that he observed between them and their environments might also be the characteristics of other individuals with similar problems.

Limitations of case studies. Case studies suffer from several limitations. Some of them are:

1. The case study data are as subjective as data collected by other researcher methods. The elements of subjectivity enter into report, particularly when judgments are made about the subject's character and motives. The researcher's personal biases and standards may influence his interpretation. Van Dalen (1973, p. 210) suggests that facts of a case study must be reported precisely and objectively and judgments must not be made until adequate evidence supports a conclusion. The researcher must exercise every possible precaution to detect data that are the product of "faulty perception, deliberate

deception, a poor memory, unconscious biases, or reporter's or subject's desire to present the right answer". He should avoid the tendency to over emphasize unusual events or to distort them for dramatic effect.

2. Although case study method attempts to examine an individual in depth, it inevitably lacks breadth. The facts about an individual or social unit may bear little relationship to other cases. Most case studies generally arise out of counselling or remedial efforts and provide information of exceptional rather than representative individuals.

3. It is impossible to either confirm or refute through empirical study the findings and results of a particular case study.

4. A worthwhile case study can rarely be completed by a single individual. It usually requires a team of experts which would include psychologists, physicians, social workers, teachers, parents, and the researcher.

2. *Causal-comparative studies.* In some investigations, the researcher attempts to explore not only what a phenomenon is like, but how and why it occurs. In such cases, the aim of the researcher is to compare the likeness and differences among phenomena to discover what factors or circumstances seem to accompany or contribute to the occurrence of certain events, conditions or practices.

Nature of causal-comparative studies. The causal-comparative studies are based upon John Stuart Mill's method of discovering causal relationships. Mill's Method of Agreement states that, "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomena" (Mill, 1846, p. 224). The following example given by Van Dalen (1973, pp. 212-213) provides a concrete illustration of this principle:

> Suppose that six students A, B, C, D, E, and F—went to a debating competition, and subsequently, three of them A, B, and C—beacme ill. To find out the cause of their illness, the doctor questioned A, B and C about what food they had eaten and discovered that every item they had selected differed expect for strawbery cream pie. In other words, the only food eaten by all those who became ill was the pie. In conformity with the Method of Agreement, the doctor concluded that the illness was caused by pie, for it was this, and only this, item on the menu that the ill students had experienced in common.

To confirm his diagnosis, the doctor could have employed Mill's Joint Method of Agreement and Differences which rules that, "If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two sets of instance differ, is the effect, or cause, or necessary part of the cause, of the phenomenon" (Mill, 1846, p. 229). To make use of this rule, the doctor would have had to find out that pie was the only food that all the ill students consumed. Furthermore, he would have had to question members of the group who did not became ill—D, E, and F. If he found the only item common to those who suffered no ill effects was the absence of pie, he would have confirmed his conclusion that Strawberry cream pie was causally related to the illness of the students.

Causal-comparative studies are employed when a researcher cannot manipulate the independent variable and establish the controls that are required in experiments. In a laboratory experiment, the experimenter controls all variables except the independent variable or variables which he manipulates to see what happens. But because of the complexity and nature of the social phenomena, a researcher while studying educational phenomena cannot always select, control, and manipulate the factors necessary to study cause-effect relationships. If a researcher, for example, wants to study emotional stability, he cannot manipulate the home background, socio-economic status, or intelligence of children and cannot place children in a situation where all factors are kept constant except one variable which is manipulated to determine what causes a particular type of emotional instability. Rather he selects children who, according to a criterion are emotionally instable and compares them with a group of emotionally stable children. After analyzing the data he may be able to identify the factors or conditions associated with the group of emotionally disturbed children and, therefore, present a possible explanation of the underlying causes of the emotional instability.

Causal-comparative method of research is useful in the situations when the experimental method is impractical or costly in time, money, and effort. In some situations ethical considerations, may prevent a researcher to use experimentation as a method of investigation. Van Dalen (1973, p. 211) states:

> Respect for living things, prevents an investigator from inflicting unnecessary pain, hardship, or harm on others, or from interfering in any way with the normal growth and development of an individual.

He has quoted some studies where the researcher searched for likeness and differences among their subjects to obtain clues about what might cause or contribute to the occurrence of a particular phenomenon.

Limitations of causal-comparative studies. Causal-comparative studies suffer from some limitations:

1. Lack of control is the serious limitation of the causal-comparative method of research.

2. It is usually difficult to identify the relevant factors causing a particular condition or phenomenon.

3. The joint method of agreement or disagreement requires that a single critical factor must be the cause for the occurrence or non-occurrence of the phenomenon. However, this condition is rarely found when a researcher is working with complex social phenomena and where the events usually have multiple rather than single causes. Furthermore, a phenomenon may result not only from multiple causes but also from one cause in one instance and from another cause in another instance.

4. When a relationship between variables is established, it is difficult to determine which is the cause and which is the effect.

5. The classification of subjects into dichotomous groups for the purpose of comparison also presents problems.

6. In comparative studies of natural situations, the researcher does not have the same control over the selection of subjects as he has in experimental studies. It is difficult to identify existing groups of subjects who are alike in all respects except for their exposure to one variable.

Though causal-comparative studies have many limitations, and they often do not produce the precise and reliable knowledge that can be gained through rigorous experimental studies, they provide means of tackling problems that cannot be probed in laboratory situations. Furthermore, they yield valuable information and clues concerning the nature of phenomena and are admirably suited to many types of field studies seeking to establish causal relationships.

3. Correlation and prediction studies. Correlation studies are a frequently used types of descriptive research concerned with determining the extent of relationship existing between variables. They are used to obtain description of existing phenomenon and enable a researcher to ascertain the extent to which variations in one variable are associated with variations in another. The magnitude of the relationship is determined through the use of the coefficient of correlation. For instance, on the basis of earlier studies a researcher may hypothesize that there is a relationship between performance on an intelligence test and a test of achievement in arithmetic. The correlation technique will help him to test his hypothesis about the relationship between these two variables as well as to assess the magnitude of the relationship.

The correlation study is relatively easy to design and conduct. It involves the collection of two or more sets of measurement on a sample of subjects and computation of the coefficient of correlation between these sets of measurements. For example, a researcher may want to test the hypothesis concerning a relationship between mathematics and physics, he would obtain the mathematics marks and the physics marks for a sample of students, say tenth grade, and then compute the coefficient of correlation between these sets of scores.

Several types of relationships can hold between the two sets of measurements. The direction of the relationship may be positive or negative; the degree of relationship between the variables may vary from perfect, to high, to average, to no relationship; the relationship may be linear or curvilinear.

In general, the magnitude of the correlation depends on the extent to which an increase or decrease in one variable is accompanied by an increase or decrease in the other—whether in the same direction (positive) or the opposite direction (negative). A high positive correlation exists if the subjects who score high (or low) on one variable tend to score similarly on the other variable. On the other hand, a high negative correlation exists if the subjects who score high (or low) on one variable tend to score opposite on the other. It may be noted that negative correlation does not indicate a lack of relationship and it is as important as a positive correlation. A perfect positive or negative correlation is rarely found in practical situations.

The examination of a scattergram may help a researcher to obtain some idea about the direction and the degree of the relationship between the variables. Figure 11.1 (*a*), 11.1 (*b*), 11.1 (*c*) and 11.1 (*d*) are the illustrations zero, positive, perfect and negative correlations between the two sets of measurements X and Y.

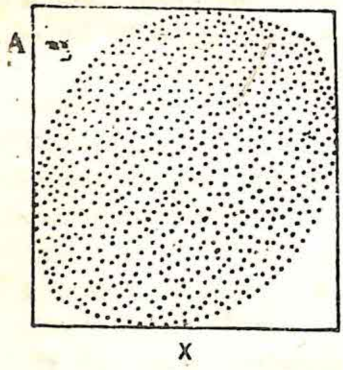


FIG 11.1 (*a*). No correlation.

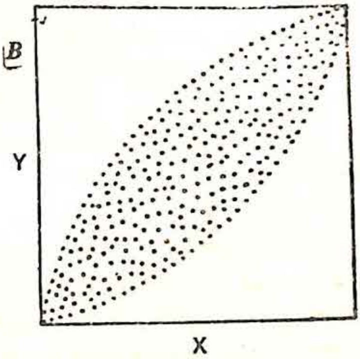


FIG. 11.1 (*b*). Positive Linear correlation.

The direction of the pattern of points, representing plotted scores, differs for a positive and negative correlation. The degree of the magnitude of the relationship between the variables is determined by examining the "Scatter" of the points. In a perfect correlation the points on the scattergram must fall in a straight line as they do in Figure 11.1 (*c*).

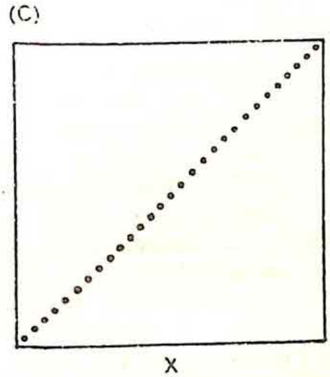


FIG. 11.1 (*c*). Perfect correlation.

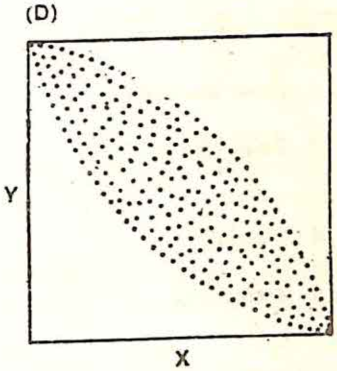


FIG. 11.1 (*d*). Negative Linear correlation.

A scattergram may also indicate whether the relationship between two variables is linear or non-linear. The pattern of points on the scattergram is dispersed around a straight line as is illustrated in the Figures 11.1 (*b*) and 11.1 (*d*). If a curvilinear or non-linear relationship exists, the points on the scattergram are dispersed around a curved line as they are in the Figure 11.1 (*e*).

The curvilinear relationships are often found when motivational, physiological, and adjustment measures are correlated with one another or with cognitive measures. A curvilinear relationship will be obtained, for example, if we correlate intelligence scores with chronological age over a life span, for intelligence increases rapidly, then reaches a plateau, and eventually declines, while age steadily increases.

Types of correlation analyses. A number of correlational techniques are employed for use with various types of data. These techniques were discussed in detail in Chapter 7. However, for the sake of recapitulation, Table 11.1 summarizes the types

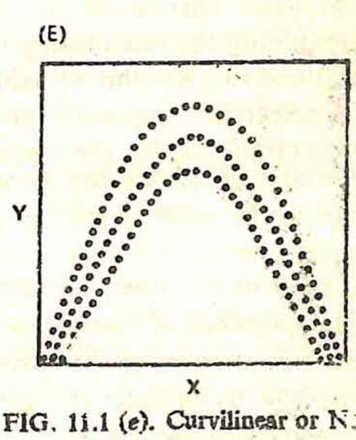


FIG. 11.1 (*e*). Curvilinear or Non-linear correlation.

of correlation coefficients used when two variables are being correlated and the types of scales for which they are appropriate.

TABLE 11.1. Types of Correlation Coefficients and Corresponding Types of Scales

Correlation coefficient	*Type of Scale*
1. Pearson product moment	1. The characteristics of both variables are expressed in interval scale.
2. Spearman rank	2. The characteristics of both variables are expressed in ordinal scale.
3. Point biserial	3. One variable on interval scale; the other a genuine dichotomous variable on a nominal scale.
4. Biserial	4. One variable on interval scale; the other an artificial dichotomy.
5. Tetrachoric	5. The characteristic of both variables are expressed in nominal scale, artificial dichotomy used with both variables and both variables have underlying continuous distributions.
6. Phi-Coefficient	6. The characteristics of both variables are expressed in nominal scale and genuine dichotomy used with both variables.
7. Contingency Coefficient	7. Both the variables are classified into two or more categories and the characteristic of both variables are expressed in nominal scale.

When a curvilinear relationship exists, a correlation ratio or coefficient of non-linear relationship (η) is computed.

Multiple correlation (R) is used if the researcher is interested in determining the relationship between one variable (independent variable) and two or more variables (dependent variables).

A researcher may not confine his attention solely to the variables he is correlating, for the magnitude of correlation between two variables may be misleading because of their relationships with other variable. In such cases, he may obtain the measure of relationship through the use of partial correlation, which eliminates or removes the effect of the other variables.

Interpretation of correlation study. The correlation study provides the result in terms of the correlation coefficient. The researcher must know how to interpret this coefficient once it is available. It may be approached in two ways: (1) by considering the degree of magnitude of a relationship, and (2) by considering the statistical significance of the magnitude of the relationship.

The degree of the magnitude of the relationship is inferred from the numerical value of the correlation coefficient, which ranges from −1.00 to +1.00 through zero. The values near zero indicate a weak relationship between the variables whereas values closer to either +1.00 or −1.00 indicate a stronger relationship. For example, suppose a researcher correlates the scores on test of physics of a group of high school students with their scores on a test of algebra and finds $r=+0.77$. He then correlates their scores on a test of English with the scores of algebra and finds $r=+0.45$. The researcher would conclude that for this group there is a stronger relationship between physics scores and algebra scores than between English scores and algebra scores.

The researcher must consider the statistical significance of the correlation coefficient when he wants to make inferences about the population from which he has drawn the sample. For this he has to test the statistical significance of the coefficient of correlation obtained from a sample. A statistically significant correlation indicates whether the obtained coefficient is different from zero at a given level of confidence. It represents evidence of an actual relationship rather than one due simply to chance.

It may be mentioned that with a reasonably large number of cases in the sample, a coefficient of correlation may be low in magnitude and yet statistically significant. For example, in a study with a sample of 400 high school boys, the correlation between the two variable was found to be +0.20. Since the sample size was large, the correlation though low, was statistically significant at .01 level. From this result of the study, it may be interpreted that in the parent population that is represented by the sample of 400 boys, the correlation is not likely to be zero. Furthermore, since it is the value of the correlation that indicates the degree of relationship between the variables, a low correlation of 0.20 indicates a low relationship between the variables even when it is statistically significant.

Moreover, for a sample of small size a coefficient of correlation may be high in value and yet it may not be statistically significant. For example, the correlation as high as +0.71 between two variable for a sample size of 8 is not statistically significant. This finding indicates that in the parent population that was represented by the sample, the correlation is likely to be zero. However, the high observed correlation suggests that there is a high relationship between the variables even when the correlation is not statistically significant.

The Table K of the Appendix provides the size of Product moment correlation r required for rejecting the null hypothesis at various levels of confidence for given sample sizes. In addition, the procedures for testing the significance of correlations have been discussed earlier in Chapter 8.

The square of the coefficient of correlation (r^2) between the variables provides another useful basis for interpretation of correlation. It shows the proportion of variance in one variable that can be attributed to its linear relationship with the other variable. In other words, 'r^2' indicates how much of their variance the two distributions have in common. For example, if scores on the tests of Physics and Chemistry show a correlation coefficient of 0.70, then $(.70)^2$ or 0.49 of the variation shown by the scores in chemistry can be attributed to the variation in the scores in physics. When the correlation coefficient between the two variables is zero, then none of the variation in one variable can be attributed to a linear relationship with the other variable.

Ary *et al.* (1972, p. 302) have pointed that a researcher must consider the following points when interpreting the coefficient of correlation:

1. A coefficient of correlation is a simple number and it should not be interpreted as percentage. An r of zero indicates no linear relationship between variables; but an r of 0.60 does not mean 60 per cent relationship between the variables. Furthermore, the r of 0.70 does not indicate twice as much relationship as that shown by an r of 0.35. It actually indicates more than twice the relationship shown by an r of 0.35. Since the coefficient of correlation ranges from —1.00 to +1.00, a difference in the absolute value of coefficients becomes more important than the same numerical difference between lower correlations would be. For example, a difference between 0.91 and 0.96 is much more important than the difference between $r=0.35$ and $r=$ 0.40.

2. A correlation coefficient gives a quantitative determination of the degree of relationship between two variables and it does not necessarily indicate a cause-and-effect relationship between them. The correlation must not be interpreted to mean that one variable is the cause of high or low score in the other variable. Frequently there are other factors that influence both of the variables under consideration. Guilford (1965, p. 105) too is of the view that "acorrelation is always relative to the situation under which it is obtained, and its size does not represent any absolute natural fact." It is relative to the circumstances under which it is obtained and should be interpreted in the light of those circumstances. To speak of the correlation between

anxiety and achievement does not convey any meaning. One needs to say which anxiety, measured under what circumstances, in what population, and to say what kind of achievement, measured by what tools, or judged by what standards.

Prediction studies. Correlation studies are particularly useful in making predictions. If we know that there is a substantial correlation between two variables, then it is possible to predict one variable from the other. In educational situations, sometimes data are available in nominal measurement rather than in any other highest type of measurements. In such situations the researcher assigns attributes to the individual cases and not any quantitative measurement on a linear scale. For example, individuals may be identified on the basis of sex,. race, caste, or political affiliation. Although no scale positions can be allocated to such data on a linear-scale, the researcher can still make predictions from them as it can be done from other information.

Guilford (1965, pp. 356-357) has conceptualized four types of predictions in psychology and education. These are predictions of (1) attributes from other attributes, such as predicting incidence of criminality from a knowledge of sex, caste, or race; (2) attributes from measurements, such as predicting criminality from known scores on tests of ability or of other behaviour traits; (3) measurements from attributes, such as predicting probable tests scores from a knowledge of sex or socio-economic status; and (4) measurements from other measurements, such as predicting achievement test scores from known aptitude test scores. Generally 'scattergram' or 'regression equations' are used in making predictions. The procedures for making such predictions have been discussed earlier in the Chapter 7.

4. *Cross-cultural and comparative studies.* In cross-cultural and comparative studies the aim of the researcher is to make explicit comparisons of a number of societies. These studies make an important contribution to our understanding of educational and social phenomena, for they seek to demonstrate whether findings concerning human behaviour are valid for all human beings or are confined to one culture. Early examples of cross-cultural and comparatives studies include development of laws of marriage and descent, relationships between certain social institutions and stages of economic development, relationship between education and society, and constructs relating to kinship derived from evolutionary theory. However, these studies made use of the data that were collected in an uncritical and unsystematic manner. With the growth of social sciences in the late nineteenth century, the dynamic relationship between education and society was recognized. Behaviourists emphasized the importance of

many cross-cultural studies relating to education. Anthropologists studied the relationship between child-training practices in various countries and cultures, the inclucation of moral values, the development of guilt, and the frequency of occurrence of crime.

With the development of more sophisticated and statistical techniques, and the utilization of the theories and tools of other fields like anthropology, psychoanalysis, and experimental psychology, the primitive methods of obtaining an understanding of the problems of education are being supplemented by empirical and quantitative methods. The recent examples of cross-cultural studies include the relationship between the content of folk tales and agression, relationship between the achievement motivation appearing in the folk tales and need achievement, relationship between various techniques of education and development of super-ego and other manifestations of personality, relationship between the alcoholic beverages and anxiety, kinship terminology and its relation to certain phenomena, relationship between sorcery and social control, relationship between the education of the child and art forms, relationship between various techniques of education and the development of super-ego and other manifestations of personality.

A review of research in comparative education reveals that researchers have compared school-state relationships; educational reforms; the problems and causes of wastage and stagnation; the accessibility to higher education; plans for the success of compulsory education; problems of caste, colour, religion and other forms of discrimination in education. Some studies have compared educational developments in African, European, Latin American, Communist, Socialist, or Capitalist countries. A number of studies have compared, national character; teacher education programmes; educational systems on various levels; economic, political and social development and change of various countries. Some studies have compared national expenditure on education and national income and have compared indicators of human resource development and economic development. A number of reports on the comprehensive comparative studies relating to illiteracy, juvenile delinquency, and education on various levels have been published by UNESCO. These reports have proved helpful in the development of planning models that express in quantitative terms a number of educational-economic relationships within several countries to help national planners predict and control change in their educational systems and achieve their economic and educational goals.

Limitations of cross-cultural and comparative studies. Cross-cultural and comparative studies suffer from many limitations. Some of them may be enumerated as under:

1. The availability, reliability, and comparability of the data present many problems. Some people are not willing to exhibit publicly behaviour that is hidden or taboo in their cultures. The available data of different cultures or countries are sometimes incomplete or incorrect and may not be described precisely and classified in the same way in all cultures or countries.

2. Reports about different cultures or the educational systems have been written at different times by people with different competencies, interests and biases. While examining such reports, the researcher should check the professional and academic status of the reporter, his knowledge of the native language, the time which he has spent in the particular culture or country, and other related factors.

3. The use of research tools like tests, interviews or observations may also create some problems. The directions and items in the tests of intelligence, aptitudes, interests and other cognitive or non-cognitive factors may not convey the same information to children in all countries. Some terms or concepts are difficult or impossible to translate into another language. If equivalent words are found in different languages, they may not have the same connotations. In case of interviews or observations, many interviewers, informants or observers do not possess a sufficiently sophisticated command of the language and knowledge of verbal and non-verbal communication. To overcome such difficulties, Van Dalen (1973, p. 238) suggests that researchers must pretest their data gathering instruments to reduce errors and to detect communication weaknesses. They may learn the second language as they have learnt the first language, through listening and responding to a native speaker, rather than through books. The researchers may translate materials into the second language, have it translated back into the first language, and compare the results by taking help of the experts in the native language.

4. The conscious or unconscious bias of the interviewer, informant or observer may also play a significant role in selecting, interpreting, or reporting the data. The bias may also operate in formulating and testing the hypotheses. The researcher, therefore, must be cautious in selecting informants or observers. He should pay particular attention to any disagreement between informant's or observers' reports and his own observations.

5. The selection of a sample may also present problems in the interpretation of the results in cross-cultural studies. If the segment of the culture selected for study is small and the period of observation is

relatively short, the researcher cannot claim that the findings would necessary apply to the whole culture. If a researcher selects a heterogeneous community comprising of two or more castes or religious groups for study, the findings may not represent the practices of many people in the group. Sometimes in a small homogeneous community or homogeneous segment of a community, a typical individual is more likely to represent the majority of the population. To overcome these problems and to have greater confidence in the findings, the researcher must select a sample of units on the basis of: (*i*) the homogeneity of each unit; (*ii*) the equivalence of the units in respect to variables other than the phenomena under study; (*iii*) geographical diversity; (*iv*). cultural, language and location independence.

6. Since cross-cultural and comparative studies are time consuming, a single person cannot gather comprehensive data from several communities, cultures or countries. Such studies need research teams which are trained in collecting diverse data in accordance with the set of criteria that specify in detail what variables to observe and how to measure them.

Developmental Studies

It is important for the educators and teachers to have reliable information about physiological, intellectual, and emotional growth of children at various ages, how they differ from one another within certain age levels, and how they change as a result of certain treatments. Such types of information are useful in taking decisions about the type of curriculum, textbooks and teaching methods. Developmental studies are used for investigating the characteristics of children and the ways in which these characteristics change with growth and development. Such studies are concerned not only with the present status and interrelationships of phenomena but also with changes that take place as a function of time. Developmental studies are also called as genetic studies. The genetic or developmental studies may take three different forms—growth studies; follow-up studies; and trend studies.

A. *Growth studies.* The teacher must have knowledge of the nature and rate of changes that take place in the children in order to teach them effectively. He must have an understanding of the interrelated factors that affect the growth of children at various stages of development and how the duration, intensity and timing of such factors in the developmental period affect growth. Growth studies are used to seek knowledge of the nature and rate of changes that take place in human organism. They are helpful to study when various aspects of growth among children are first observable, accelerate further, remain rather stationary, attain optimal development, and decline.

Growth studies can be either longitudinal or cross-sectional.

Longitudinal studies. Longitudinal studies follow the same subject, a group of subjects, or an institution over a relatively long period of time. It is a kind of extended case-study as it involves a long-life span of the same subject, group or institution. A researcher, for example, may test and measure the same student on a trait of personality when he is ten, eleven, twelve, thirteen, and fourteen years of age and plot his individual growth patterns for this trait during these years. Therefore, he could assess the change in the trait over a period of time at different age levels for the student. Since the researcher is dealing with the same student, the effect of such factors like family background, etc. will remain constant and the differences he observes between two age levels may be interpreted as change in the particular personality trait related to the growth of the student. The longitudinal type of research is mostly used in clinics and laboratories. However, its use in the fields of education is not uncommon.

Some studies are especially good examples of carefully planned and executed longitudinal investigations. Terman and his co-workers (1925, 1947) made a comprehensive longitudinal investigation of gifted children. The findings of this study have made a significant contribution to the understanding of the growth of gifted children and also set the methodology for further research in this area.

Gesell (1928, 1940, 1946) studied motor and physical growth and the development of emotional expression, philosophic outlook, adaptive behaviour, language, interpersonal relationships, and personal-social behaviour of children from infancy to age 16. His work extends over half a century, from the time he entered the Yale Clinic of Child Development in 1911 to his death in 1961. These longitudinal studies constitute a major contribution to the understanding of the child and his development.

Honzik, Macfarland, and Allen (1949), using longitudinal method, studied the fluctuations of IQ scores of children from age two to age 18. They found that the IQ scores of over half the group changed 15 or more points, the scores of 9 per cent of the group changed 30 or more points, and that scores separated by a short time period correlated more highly than scores seperated by a long time period. It was also found that the children whose IQ scores exhibited consistent downward trends belonged to lower socio-economic groups and those whose scores showed consistent upward trends came predominantly from higher socio-economic groups.

Limitations of longitudinal studies. Though longitudinal method is useful in accumulating data for the same subject or subjects at various levels and helps in intensive studies of individual or individuals, it suffers from the following limitations:

1. Longitudinal studies have sampling weaknesses: (*a*) The data in these studies are usually obtained from a single individual or a limited number of individuals. If the sample selected proved to be a poor one, there is nothing that can be done to it, nor can new longitudinal variables for investigation be introduced after the completion of the study.

(*b*) When the individual are selected from a community with a stable population for the sake of cooperation and keeping track of them throughout the years of the study, a bias in the findings may get introduced because of the low mobility of the group. In some cases, it becomes difficult to maintain the cooperation and track of the individuals for a longer period.

(*c*) It becomes extremely difficult to get the data in continuity from the individuals who are selected from a mobile community.

(*d*) Longitudinal studies may give accurate descriptions of the growth of the individuals studied, but these data are not necessarily representative of the total population from which the individuals are selected.

(*e*) The individuals of a particular community who voluntarily cooperate or who drop out may differ from other members of the community. Moreover, sometimes it may be impossible to obtain complete data for all individuals over the years because some of them may die, become ill, move or lose interest in further participation. If the researcher has selected a small group at the start of the study, the size of the sample may get reduced to a point of little value at the final stage of the study.

To minimize the dropout problem, the researcher may take some precautions. He should select a group considerably larger than he reports in the final results. The researcher may also select samples from an area of low geographical mobility and try to establish good rapport with the individuals and their associates by exhibiting an interest in them and in their problems. In addition, the researcher may keep record of the cause and number of dropouts; study the variations between the individuals who cooperated throughout the investigation and the dropouts; and when possible, relocate some dropouts and obtain at least a general idea about their growth and development.

2. Another limitation of the longitudinal method is that the researcher usually cannot make improvements in his measuring tools or

instruments as his study develops without disrupting the continuity of the procedures. At the later stage of the study, if the researcher discovers new and better tools and decides to use them, the data he gathers with these tools are not likely to be comparable. Van Dalen (1973, p. 247) too has pointed out that:

> Shifts in observational procedures are often necessary because of the absence of a single technique that is applicable across the range of ages. The same instrument cannot be used to discrimate between high and low activity of new born, five-year-old, and sixteen-year-old children.

He is of the view that the researcher cannot assume the equivalence of different measuring tools and techniques, but he can take the precaution of using two or more tools that are applicable for a particular age and evaluate the relationship among them.

3. Sometimes longitudinal studies become unwieldy and unmanageable because of the wide range of behaviour to be observed, large number of the subjects in the sample, or less duration of the time period in which the study has to be completed. In such cases the study should not be conducted by one individual but a group of researchers. The group may be divided into two teams, one team may collect data and maintain contact with the subjects and the other team may tabulate, organize, analyse and interpret the data. Two teams approach reduces the chances of delay and contamination.

4. Longitudinal studies require extensive facilities, considerable amount of money and maximum perseverence on the part of research personnel over a number of years. Such studies, therefore, may be undertaken by research organizations in the form of -projects and not by individual researches.

Cross-sectional studies. Many of the limitations of the longitudinal method are overcome in the cross-sectional studies. When conducting cross-sectional study, rather than following the same group of individuals and taking their repeated measurements over a relatively long period of time, the random samples of individuals of successive ages are selected and one set of measurements of different individuals from each age level are taken as the basis for developing growth norms. The statistics derived from the samples are compared and the conclusions are drawn about the growth of individual with respect to the variables studied.

Limitations of cross-sectional studies. Though in cross-sectional studies, the researcher selects age differing groups that are as alike as possible in respect to sex, intelligence, socio-economic status, or any other variable that may significantly affect the growth factor being studied, yet they have the following limitations:

1. The researcher cannot remove all variations due to the variables other than age in which groups differ and the differences due to extraneous variables may seriously affect the results.

2. In cross-sectional studies, the measurements taken of individuals at each age level may not be comparable because the groups may differ on variables other than age. The subjects in the older group may not have had the same home or other environmental experiences as those in the younger group. The older group may exhibit a fearful behaviour that is not necessarily related to age but rather to the depression they experienced as children. Van Dalen (1973, p. 244) is of the opinion that:

> Our cultural framework is less stable than it once was; hence, the cross-cultural method may yield non-comparable data.'; Because of rapid technological developments and social changes, twenty-year-old subjects may not have had the same kind of early experiences as seven-year-old subjects.

3. Another serious limitation of cross-sectional studies is that chance differences between samples may seriously bias the results. For example, suppose a researcher may by chance draw a sample of fourth grade students who are more mature than average and a sample of third grade students who are less mature than average with the result that the difference between groups appears to be much smaller than it really is.

The longitudinal studies are generally considered more acceptable than cross-sectional studies because they have the advantage of continuity and permit the researcher to record individual fluctuations, which are frequently of greater value and interest to him than the overall growth pattern itself. However, if a researcher wants to learn the characteristics of typical children at various stages, the cross-sectional method is preferred because of the greater possibility of obtaining large samples with this technique. Moreover, the selection of taking large samples may in many cases outweigh the limitations of the cross-sectional method (Ary *et al.*, 1972, p. 295).

2. *Follow-up studies.* Follow-up studies follow the method which somewhat resembles the longitudinal method. Such studies aim at investigating the subsequent development of individual or individuals after a specified treatment or condition. In case study approach, a researcher may be interested to follow-up a case in order to determine whether the treatment is successful. If the follow-up of a case indicates that its

progress is unsatisfactory, a new diagnosis of the problem or problems is made. The techniques of experimentation are employed to evaluate the success of treatment or condition.

The studies by Terman and his associates (1947) are good examples of follow-up studies. They studies gifted subjects in 1921-22. The subjects were studied after six years and again in 1936, 1940 and 1945. After the follow-up it was found that in comparison to other children, the gifted children had better physical and mental health as adults and they matured into gifted and successful adults.

3. *Trend studies.* Trend studies are used to obtain and analyse social, economic, or political data to identify trends and to predict what is likely to take place in the future. Some educational or other organizations conduct such studies in order to plan effectively and to meet the demands that will be made on them in the future.

Trend studies are undertaken through documentary analysis or surveys at repeated intervals. Such surveys help the researchers to study the rate and direction of changes and to use these trends to predict future status. Sometimes trend study may combine historical, documentary, and survey techniques.

For the sake of illustration, a researcher may conduct or examine studies of birth rate trends, the changing age structure of the population, enrollment of children in schools, the amount and direction of population mobility in order to determine when and where to provide new educational facilities in terms of school buildings, teacher training programmes, or recruitment of teachers. State education departments may wish to study the progress that is being made in various areas or to measure the need for changes in syllabi, texbooks, guidance programmes, and working conditions of teachers and other personnel. To do this they may undertake trend studies that relate to student achievement; changes in attitudes of teachers and parents towards school programmes and policies; the types of available jobs and anticipated job opportunities; and day-to-day views of educators, politicians and newspaper editors.

STEPS IN DESCRIPTIVE RESEARCH

The process of descriptive studies is not different from other forms of research. Since such studies, describe and interpret what conditions or relationships exist at present, the researcher may adopt the following steps:

Selection of the Problem

A researcher may be concerned with conditions or relationships that exist, practices that prevail, beliefs, points of view or attitudes that are held, processes that are going on, effects that are being felt or trends that are developing, and may select the problem accordingly from the area or field in which he is interested.

Statement and Definition of the Problem

The researcher must state the problem clearly as it is done in case of other types of research. The statement must identify the variables involved in the study. It should specify clearly whether the study is merely seeking to determine the presents status of these variables or whether it will also explore relationships between the variables.

Identification of Data

After stating and defining the problem, the next step for the researcher is to list the data to be collected for the study. He has to specify whether the data are of qualitative or a quantitative nature and whether the data will be collected in the form of counts, test scores, responses to questionnaries, interviews, and so on.

Selection or Development of Tools

The nature of the data to be collected helps the researcher to select the appropriate tools for the study. If the readymade tools are not available, the researcher has to develop his own tools. Questionnaires, interviews, psychological tests, rating scales, schedules and attitude scales are the most frequently used tools for descriptive research. If the researcher uses readymade tools, he should satisfy himself about their reliability, validity, and suitability for sample chosen for the study. If the researcher develops his own tools, he should try them out with a small group in order to evaluate them and make modifications if necessary.

Selection of the Sample

The researcher must select the sample about which he wishes to seek information using appropriate sampling techniques. The sample selected should adequately represent the population.

Collection of Data

The researcher should specify the practical schedule for gathering the data from the sample selected for the study with the help of appropriate tools.

Analysis and Interpretation of Data

The data collected is quantified in the form of counts, test scores, responses to questionnaries, etc. These are analysed and interpreted with the help of appropriate parametric or non-parametric statistical tests.

Writing of the Research Report

It is the last stage in the descriptive research as in any other form of research. The researcher should exercise extreme caution in generalizing conclusions and reporting them with all the limitations of the study.

POST TEST

1. Describe the nature of descriptive research.
2. Compare the descriptive method with the historical method of educational research.
3. Describe the value of descriptive research in education.
4. List and describe the various Types of descriptive research.
5. Describe the nature of survey studies.
6. List and describe the various types of survey studies.
7. Describe the nature and purpose of school surveys.
8. List and describe the various types of school surveys.
9. Describe the nature and various types of survey testing.
10. Describe the uses of achievement testing, intelligence testing and personality testing in educational research.
11. Describe the nature and purpose of status studies.
12. Describe the nature and purpose of school appraisal.
13. Describe the nature and purpose of financial studies.
14. Describe the nature and purpose of curriculum studies.
15. Describe the nature and purpose of building surveys.
16. List and describe the various ways in which the school surveys can be organized.
17. Describe the nature and purpose of job analysis.
18. Describe the nature and purpose of 'Content' analysis.
19. Describe the nature and purpose of public opinion surveys.
20. Describe the nature and purpose of social surveys.
21. Describe the nature of interrelationships studies.
22. List and describe the various types of interrelationships studies.
23. Describe the nature of case study.
24. List and describe the steps of the case study.
25. List and describe the advantages of case studies.
26. List and describe the difficulties and limitations of case study method. What improvements would you suggest to make this method more useful in educational research?
27. Describe the nature of causal-comparative studies.
28. List and describe the uses of causal-comparative studies.

29. List and describe the limitations of causal-comparative studies.
30. Describe the nature of correlation studies.
31. List and describe the various types of correlation analyses.
32. How would you interpret the results of a correlation study?
33. Describe the nature and purpose of prediction studies.
34. Describe the nature of cross-cultural and comparative studies.
35. Describe the uses of cross-cultural and comparative studies.
36. List and describe the limitations of cross-cultural and comparative studies.
37. Describe the nature of developmental/genetic studies.
38. List and describe the various types of developmental studies.
39. Describe the nature of growth studies.
40. List and describe the two types of growth studies.
41. Describe the nature and purpose of longitudinal studies.
42. List and describe the limitations of longitudinal studies.
43. Describe the nature and purpose of cross-sectional studies.
44. List and describe the limitations of cross-sectional studies.
45. Describe the nature and purpose of follow-up studies.
46. Describe the nature and purpose of trend studies.
47. List and describe the steps which the researcher may adopt in conducting the descriptive type of research.

12

12. Experimental Method

The five sections of this chapter reflect the following purposes. The first section describes in detail the nature of experimental research. It also discusses the four essential characteristics of experimental research: control, manipulation, observation, and replication. The second section, which is quite brief, deals with the value of experimental research in education. The third section describes the steps which the researcher may adopt in conducting experimental research. The fourth section describes in detail the nature of experimental designs. It also discusses their appropriateness, adequacy of control, and validity. The fifth and final section deals with different types of experimental designs: pre-experimental designs, true experimental designs, factorial designs, quasi-experimental designs, and time-series designs.

Learning Objectives

After carefully reading this chapter, you will be able to meet the following objectives:

1. Describe the nature of experimental research.
2. List and describe the essential characteristics of experimental research.
3. Define the term control.
4. List and describe the purposes of control in an experiment.
5. List and describe the methods for controlling intersubject differences.
6. List and describe the methods for controlling situational variables.
7. Define and illustrate the terms manipulation, observation and replication.
8. Describe the value of experimental research in education.
9. List and describe the steps which the researcher may adopt in conducting the experimental type of research.
10. Define an experimental design.
11. List and describe the important criteria which the researcher must keep in mind while selecting an experimental design for conducting his experiment.
12. Define validity of the design.
13. List and describe two general types of validity of the design.
14. Define internal validity of the design.
15. List and describe the extraneous variables which affect significantly the internal validity of the design.
16. Define external validity of the design.
17. List and describe the types of external validity of the design.
18. Define population validity of the design.
19. Define ecological validity of the design.
20. List and describe the factors which contribute towards achieving ecological validity of the design.

21. Define the terms: experimenter, subject, independent variable, dependent variable, control group, experimental group, pretest and posttest.
22. List and describe various types of experimental designs.
23. Describe the nature of pre-experimental designs.
24. List the various types of pre-experimental designs.
25. Describe the nature of One-Group Pretest-Posttest Design.
26. Illustrate the use of One-Group Pretest-Posttest Design.
27. List and describe the limitations of One-Group Pretest-Posttest Design.
28. Describe the nature of Two-Groups, Static-Design.
29. Illustrate the use of Two-Groups, Static Design.
30. List and describe the limitations of Two-Groups, Static-Design.
31. Describe the nature of True Experimental Designs.
32. List the various types of True Experimental Designs.
33. Describe the nature of Two-Groups, Randomized Subjects, Posttest-only Design.
34. Illustrate the use of Two-Groups, Randomized Subjects, Posttest-only Design.
35. List and describe the advantages of Two-Groups, Randomized Subjects, Posttest-only Design.
36. List and describe the limitations of Two Groups, Randomized Subjects, Posttest-only Design.
37. Describe the nature of Two-Groups, Randomized Matched Subjects, Posttest-only Design.
38. Illustrate the use of Two-Groups, Randomized Matched Subjects, Posttest-only Design.
39. List and describe the advantages of Two-Groups, Randomized Matched Subjects, Posttest-only Design.
40. List and describe the limitations of Two-Groups, Randomized Matched Subjects, Posttest-only Design.
41. Describe the nature of Randomized Groups, Pretest-Posttest Design.
42. Illustrate the use of Randomized Groups, Pretest-Posttest Design.
43. List and describe the advantages of Randomized Groups, Pretest-Posttest Design.
44. List and describe the limitations of Randomized Groups, Pretest-Posttest Design.
45. Describe the nature of Randomized Solomon Three-Group Design.
46. Illustrate the use of Randomized Solomon Three-Group Design.
47. List and describe the advantages of Randomized Solomon Three-Group Design.
48. List and describe the limitations of Randomized Solomon Three-Group Design.
49. Describe the nature of Randomized Solomon Four-Group Design.
50. Illustrate the use of Randomized Solomon Four-Group Design.
51. List and describe the advantages of Randomized Solomon Four-Group Design.
52. List and describe the limitations of Randomized Solomon Four-Group Design.
53. Describe the nature of Factorial Designs.
54. Describe the nature of Simple Factorial Design of 2 by 2 (2×2).

55. Illustrate the use of Simple Factorial Design of 2 by 2 (2×2).
56. List and describe the advantages of Factorial Designs.
57. List and describe the limitations of Factorial Designs.
58. Describe the nature of Quasi-Experimental Designs.
59. List the various types of Quasi-Experimental Designs.
60. Describe the nature of Non-randomized Control Group, Pretest-Posttest Design.
61. Illustrate the use of Non-randomized Control Group, Pretest-Posttest Design.
62. List and describe the advantages of Non-randomized Control Group, Pretest-Posttest Design.
63. List and describe the limitations of Non-randomized Control Group, Pretest-Posttest Design.
64. Describe the nature of counterbalanced Design.
65. Illustrate the use of counterbalanced Design.
66. List and describe the advantages of counterbalanced Design.
67. List and describe the limitations of Counterbalanced Design.
68. Describe the nature of Time-Series Designs.
69. List the various types of Time-Series Designs.
70. Describe the nature of One-Group Time Series Design.
71. Illustrate the use of One-Group Time-Series Design.
72. List and describe the advantages of One-Group Time-Series Design.
73. List and describe the limitations of One-Group Time-Series Design.
74. Describe the nature of Control-Group Time-Series Design.
75. Illustrate the use of Control-Group Time-Series Design.
76. List and describe the advantages of control-Group Time-Series Design.
77. List and describe the limitations of Control-Group Time-Series Design.

The two preceding chapters dealt with historical and descriptive research methods. The next broad category of educational research involves experimentation. It differs from descriptive studies, in that the researcher has some degree of control over the variables involved and the conditions under which the variables are observed. The researcher deliberately manipulates some aspect of the experiment in which he is interested. He causes certain things to happen, and he observes how the condition is affected or changed. Although experimentation is the classic laboratory method of psychology, physics, chemistry, biology and other sciences, it has been effectively used in non-laboratory educational settings such as the classroom.

NATURE OF EXPERIMENTAL RESEARCH

In descriptive research, a researcher may analyse the data he has collected and discovers that different variables are related. The relationships between the variables can be interpreted in several ways. The researcher interprets his findings in one way and faces difficulty in eliminating plausible alternate explanations of his findings because of the lesser control over the factors studied. Experimental method, on the other hand, provides for much control and, therefore, establishes a systematic and logical association between manipulated factors and observed effects. The researcher defines a problem and proposes a tentative answer or hypothesis. He tests the hypothesis and accepts or rejects it in the light of the controlled variable relationship that he has observed.

The earliest assumptions of experimental method are based upon the law of the single variable. John Stuart Mill (1846) defined this law in his work *Methods of Experimental Inquiry*. He stated five canons or rules of experimental research: (1) the method of agreement, (2) the method of difference, (3) the joint method, (4) the method of residues, and (5) the method of concomitant variations. These rules served as useful guides in the design of early experiments and are now used chiefly as guides in the planning of experiments.

The human events are complex and rarely are they the result of restricted relationship between single variables. They are usually the consequence of the interaction of many variables. Since the law of single variable assumes a highly artificial and restricted relationship between single variables, it failed to provide an adequate approach to experimentation for the study of human beings in the laboratory. Moreover, for various reasons it was not always possible or convenient to study human beings in the laboratory. Certain non-laboratory techniques were, therefore, devised for observing human beings experimentally in groups outside the laboratory. The contributions of R.A. Fisher in terms of his concept of achieving pre-experimental equation of conditions through random selection of subjects and random assignment of treatments have provided an effective and sound method of conducting-realistic experiments with human beings. His techniques of analyses of variance and co-variance made it possible to study complex interactions through factorial designs.

There are four essential characteristics of experimental research: (1) Control. (2) Manipulation, (3) Observation, and (4) Replication.

Control

Control is the essential ingredient of experimental method. It refers to the extent to which different factors in an experiment are accounted for. Since more of the factors are accounted for with accuracy and more control is being enforced, the researcher has more confidence that his results are dependable.

Van Dalen (1973, pp. 263-264) has pointed out that in an experi ment, the researcher seeks to control variables for the following purposes:

1. **Achieving Isolation**
   To prevent a factor other than the independent variable from affecting the dependent variable, the researcher may remove the unwanted or interfering variable, or he may keep constant its effect or equalize its presence in the experimental and control groups.

2. **Achieving Changes in Magnitude**
   A researcher may strive not only to isolate the independent variable but also to ascertain how much effect it contributes. To achieve this objective, he must be able to vary the magnitude of the experimental variable.

3. **Achieving Quantitative Evaluation**
   The ultimate goal of a researcher is to express the magnitude of the variable in quantitative terms. He wants to know not merely that one expression of a variable is larger or smaller than another but precisely how much larger or smaller it is. If two variables are functionally related, he wants to state not merely that they are positively or negatively related but rather the specific degree of relationship in terms of some numerical value.

High degrees of control is much easier to achieve in the laboratory setting than in the situations outside the laboratory. In the laboratory, the researcher deals with a limited number of factors and he can manipulate the conditions at will. He can be sure of the changes that have taken place and within a limited amount of time he can measure the effects with great precision.

Methods of Control

As discussed earlier, the main purpose of control in an experiment is to arrange a situation in which the effect of variables can be measured. Since human nature is complex, there are many variables present while dealing with human beings in an experimental setting. The attempt of a researcher to reduce educational problems to the operation of a single variable is, therefore, not only unrealistic but impossible.

For the sake of illustration, suppose a researcher wishes to study the effectiveness of conventional method and programmed instruction in teaching of algebra. For this he has to frame two groups of subjects who are identical in every respect except the way in which they are taught algebra. Since it is impossible to have two absolutely identical groups of subjects, the researcher will try to establish two groups that are as identical as possible in respect to those variables that are related to the achievement in algebra. For this he may select the variables of general intelligence, aptitude in mathematics, achievement motivation, and the like. He can easily ignore the variables that are highly unrelated to achievement in algebra. Since the law of a single variable cannot be applied absolutely, the researcher will seek to follow it as strictly as possible in all relevant variables.

In experimental studies in education the researcher has, therefore, to direct his effort towards controlling the variables which are significantly related to the dependent variable. Such variables are responsible for any relevant pre-existing differences between the subjects used in the experiment. In addition to inter-subject differences, there are some situational variables that might operate in the experimental situation itself. If the relevant and situational variables are not controlled in an experiment, the researcher cannot be sure whether these are the independent variables or the incidental differences operating in the groups that are producing the difference in the dependent variable.

Various methods which are commonly used to control inter-subject differences and situational variables are discussed as under.

Methods for Controlling Inter-subject Differences

Following procedures are commonly used to minimize inter-subject differences so as to increase equivalence among groups that are to be subjected to various experimental conditions.

1. *Random assignment of svbjects to groups*. Randomization is based on probability theory. It is considered the best technique of attaining experimental equivalency.

According to Kerlinger (1978, pp. 123-124):

> Randomisation is the assignment of objects (subjects, treatments, groups) of a universe in such a way that, for any given assignment to a subset, every member of the universe has an equal probability of being chosen for that assignment . . . . Since, in random procedures, every member of a population has an equal chance of being selected, members with certain distinguishing characteristics – male or female, high or low intelligence, Republican or Democrat, dogmatic or not dogmatic, and so on and on – will, if selected, probably be counterbalanced in the long-run by the selection of other members of the population with 'opposite quantity or quality of the characteristic.

Randomization is not a haphazard or arbitrary method of assigning objects of a universe to subsets; it is a scientific and systematic procedure.

In the preceding illustration, suppose a researcher has available, a group of subjects from which he has to select some number for experimentation. The selected number of subjects will be divided into two groups for different treatments and comparison. In assigning subjects to two groups, the researcher may use the method of randomization to avoid the introduction of biases due to personal judgement and of the characteristics of the subjects themselves.

To obtain two random groups, one to be taught through conventional method and the other through programmed instruction, the researcher could give a serial number to all of the available subjects and employ tables of random numbers, or use some other known random procedure to select needed number of subjects for the two experimental treatments. Similarly, he could assign teachers. classrooms, equipment, and class periods strictly on a random basis. Then with the help of appropriate statistical tests, the researcher would decide whether the observed difference in the achievement in algebra is sufficiently larger than what would be expected from sampling error. It may be noted that in assigning subjects randomly to two groups, one can assume that the groups are comparable at the start of the experiment on the dependent variable as well as on the extraneous variable. Ary *et al* (1972, p. 222) have pointed out that through randomization of subjects not only these extraneous variables, that a researcher can identify, are randomized out but also other relevant extraneous variables, which are unknown to him, can be expected to randomize out.

2. *Matching subjects with random assignments.* Another method that is used for assigning subjects to groups is to match individual subjects on as many extraneous variables as a researcher can identify that might affect dependent variable and then apply some random technique to assign one member of each matched pair to the experimental conditions. The researcher, for example, in the preceding illustration may match the subjects on the pre-achievement scores in algebra and/or any other variables that are known to have an effect on the dependent variable, such as scores on an intelligence test or an aptitude test in mathematics, achievement motivation test, or socio-economic background. Then one member of each pair is randomly assigned to the two groups.

The researcher may be encountered with several difficulties. The first and the important of these is to determine what variable or variables are significantly related to the dependent variable and which of these to use as a basis for matching. Variables such as intelligence, socio-economic status, sex, age, achievement motivation, pretest scores on the dependent variable, etc. are commonly used as the basis for matching. The experts are of the opinion that the variables on which subjects are matched must correlate as high as 0.50 or more to the dependent variable. To quote Ary *et al.* (1972, p. 223):

> As a general rule, we suggest that unless the variable correlates 0.50 or higher with the dependent variable it should not be employed for matching procedure .since it would do little to increase the precision of the study. Ideally we would like to match on two or more variables that correlate well with the dependent variable and do not correlate significantly with each other. However, when we try to match on more than two variables it becomes almost impossible to find subjects who are well matched on these several variables. Thus, we lose subjects because no match can be found for them.

Another problem that may be faced by the researcher is how closely to match the subjects on the variable or variables. If he matches the subjects closely, it increases the precision of the method at the cost of reducing the size of sample and increasing sampling bias into the study.

Generally three matching procedures are used in experimental studies. Keeping in view his situation, the researcher must decide what matching procedure is feasible in a particular situation.

(*i*) *Procedure of subject-to-subject matching*. In subject-to-subject matching, the researcher seeks to locate two members from the available subjects who score within the limits decided upon by him. For the sake of illustration, suppose the researcher selects socio-economic background as the matching variable, then he may identify two subjects who are within 3 points of each other on the socio-economic status scale, and then randomly assigns one member of the pair to one group and the other member to the other group. It is always possible to match subjects on one variable. However, if there are other two or more relevant variables, then it becomes extremely difficult for the researcher to find pairs who match on all such variables and, therefore, most of the subjects are not selected for the study.

(*ii*) *Matching for mean and standard deviation*. To overcome the problems of subject-to-subject matching, sometimes it is worthwhile to match groups rather than individuals on the relevant variable or variables. The researcher seeks to show that the two groups do not differ significantly in terms of mean and standard deviation on the matching variable or variables. For example, in an experimental study, intelligence, socio-economic status and achievement motivation are considered the relevant matching variables. In such a study, the researcher may analyse the scores on intelligence test, socio-economic status scale and achievement motivation test, and try to find out that there is no significant difference in the means and standard deviations of the scores on the selected tests. He than randomly assigns the groups to two experimental conditions. Some experts are of the opinion that matching on the basis of groups is less precise in comparison to individuals matching. Moreover, the researcher may feel the problem in identifying the groups that match on all of the variables that may be correlated with the dependent variable.

(*iii*) *Ranking of subjects on the matching variable.* A third method of matching is to place all the available subjects in rank order on the basis of their scores on the matching variable. Regardless of the actual difference, the first two subjects are selected from the rank order list and these constitute the first pair. One subject of this pair is then randomly assigned to one of the groups and the other to the second group. Similarly, the next two subjects on the rank order list are chosen and again one is randomly assigned to the first group and the other to the second group. This procedure is continued until the whole list of subjects is exhausted. This method is also less precise than that of the subject-to-subject matching.

3. *Random assignment on the basis of homogeneous selection.* Another procedure that is used to make groups comparable on an extraneous variable is to select groups that are as homogeneous as possible on that variable. For example, if a researcher is confident that sex is a variable that might affect the dependent variable, then he could select subjects of a particular sex. By selecting only male students, for example, the researcher would control the effects of sex as an extraneous variable. Similarly, if socio-economic status is likely to be a variable that may affect the dependent variable of the study, the researcher could select subjects within a restricted range of socio-economic status. By this method the researcher is able to control the effects of socio-economic status. After selecting the subjects from the homogeneous population, the researcher could randomly assign subjects to two groups and be confident that the groups were comparable on the relevant variable. Although the selection of subjects from the homogeneous group is useful in eliminating the problems of subject-to-subject matching, it has the disadvantage of decreasing the extent to which the results can be generalized beyond some situations. If a researcher studied teaching behaviour of a group of inservice teachers of a particular age-range or sex, he could not generalize from the findings of this study the teaching behaviour of inservice teachers of other age or sex. For such generalizations he would have to repeat studies with teachers from other sex or different age ranges.

4. *Technique of analysis of co-variance.* Because of the problems of the matching procedures discussed earlier, the technique of analysis of co-variance is used to control the variation within the groups. This technique analyses the differences between the two groups, subjected to the two different experimental treatments, on the dependent variable after taking into account any initial difference between the groups on

pretest measures or any other relevant independent variables.

In the analysis of co-variance, one or more co-variates in addition to the dependent variables are measured. A co-variate is a variable that the researcher has not been able to control in the experiment and is believed to affect the dependent variable. The technique of analysis of co-variance is not limited to the use of only one co-variate.

A detailed discussion on the use of analysis of co-variance in the experimental research has been provided in Chapter eight.

5. *Method of using subjects as their own controls.* Another method of control is to assign the same subjects to two experimental treatments and then to obtain measurements of the subjects first under one treatment and then under the other. Although this method is an efficient method of control, it is not feasible in certain circumstances. A researcher, for example, is interested to know the difference in learning time between two different lists of nonsense syllables—one list with high association value and the other with low association value—in an experiment on retention. In such an experiment, the researcher can find the difference in learning time between two lists for each subject and then test the average difference in learning time for all subjects with the help of an appropriate statistical test for significance. But the effects of relative ease of learning the two lists, fatigue, and other interference effects cannot be partialled out completely and no reliable conclusions could be drawn. Moreover, in some studies the researcher while exposing the subjects to one experimental condition could not use them for the other experimental condition. In teaching some material with two different methods, for example, a researcher cannot teach them with one method and then make them to forget what they learnt and teach them with another method.

Methods for Controlling Situational Variables

There are three methods commonly used to control situational variables.

1. *Method of holding situational variables constant.* In the method of holding situational variables constant, the researcher treats all the available subjects alike except for their exposure to the independent variable. For example, in a reading experiment the researcher may take equal number of subjects in two groups, teach them by the same teacher, use the same instructions, apparatus, and tests. The groups may be taught in the same classroom and at the same time of the day, and in the same environmental conditions such as temperature, presence or absence of distracting noise, furniture in the room and the like.

2. *Method of randomization.* If the situational conditions cannot

be treated alike, the researcher should try to balance them by randomization. Suppose in the study mentioned above it is not possible to have the same teacher for both the groups. Then the researcher may divide the two major groups into two smaller sub-groups and randomly assigns half subjects of the first major group and half subjects of the second major group to each teacher. The same procedure could be used to randomize other situational variables like time, apparatus and tests.

3. *Method of manipulating situational variables.* The situational variables can also be controlled by manipulating them systematically. In some educational experiments, the researcher can use a sequence of experimental and control conditions in order to control what Ary [illegible] *al.* (1972, p. 227) have called 'progressive effects', like those of practice and fatigue. This, according to them can be done by "controlling the order in which experimental conditions are presented through a counter-balancing; half the subjects may receive an *AB* order and the other half a *BA* order." The authors are of the opinion that this method not only controls the potentially contaminating effect of order, but can also help in providing an estimate of the size of the order effect by finding whether the means of *A* and *B* values obtained in the two sequences are different.

Manipulation

Manipulation of a variable is another distinguishing charactaristic of experimental research. It refers to a deliberate operation of the conditions by the researcher. In contrast to descriptive research, in which the researcher simply observes conditions as they occur naturally, the researcher in the experimetal research actually sets the stage for the occurrence of the factor whose performance is to be studied under conditions in which all other factors which might complicate the observation are controlled or eliminated. In the process of manipulation, a predetermined set of varied conditions are imposed on the subjects selected for the experiment. The set of varied conditions is referred to as the independent variable, the experimental variable, or the treatment variable. The independent variable may assume two or more values and the differences in the values may be of qualitative or quantitative nature. Sex, methods of teaching, attitudes, socio-economic status, classroom environment, personality characteristics, types of motivation, etc. are some common examples of independent variables in educational research.

If the researcher compares the performance of male and female students, sex is the independent variable. If he compares two methods of teaching, method of d :teaching is the independent variable and can be manipulated by the teacher. Since sex in the former case cannot be manipulated, it is called an 'assigned independent variable'. The researcher may manipulate a single variable or a number of variables simultaneously. He must get an expertise in the technique of 'multivariate analysis' so as to save time and energy in the simultaneous investigation of a number of variables studied singly and in interaction.

Observation

In the experimental research, the researcher studies the effect of the manipulation of the independent variable on a dependent variable. The dependent variable, for example, may be learning some task. Since learning cannot be measured directly, the researcher can only estimate it through such measures as scores on a test. Strictly speaking dependent variable, therefore, is scores on a test or observations with respect to some characteristics of the behaviour of the subjects used in the experiment.

Replication

No matter how objectively and carefully a researcher attempts to control the extraneous variables through the methods of randomization or other methods, still some discrepancies invariably remain and influence the results of the experiment. The researcher can take care of such discrepancies through the replication of the study.

Replication is a matter of conducting a number of sub-experiments within the framework of an overall experimental design. The researcher, instead of comparing a single control case with a single experimental case, makes a multiple comparison of a number of cases of the control group and a number of cases of the experimental group, all within the same experimental framework. He assigns equivalent subjects to each of the control and the experimental groups at random and thus considers the comparison of each pair as an "experiment" in itself (Mouly., 1963, p. 337). If a researcher, for example, conducts an experiment involving 30 subjects in each of the experimental and control groups; he is actually conducting thirty parallel experiments in one. In some experimental situations, a number of control and experimental groups, each consisting of equivalent subjects assigned at random to one or another of the two groups are combined within the design of a single experiment.

VALUE OF EXPERIMENTAL RESEARCH IN EDUCATION

The experimental research is not considered a precise method of research in the field of education because of the complex nature of the human beings and problems of controlling the extraneous variables. However, in spite of all such difficulties, experimentation has been put to various uses in solving educational problems. Campbell and Stanley (1963) are of the opinion that:

> The experiment is the only means for settling disputes regarding educationa practice, the only way of varifying educational improvements, and the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favour of inferior novelties.

Experimental research is used to determine and evaluate the adequacy and effectiveness of the educational and instructional objectives through the measurement of their outcomes. After evaluating the efficacy of objectives, the suggestions are made for the formulation, execution and modification of educational programmes and classroom practices. The classroom teacher uses experimentation to evaluate the effectiveness of certain learning experiences, planned and organized, to achieve some desired objectives. Effectiveness of teaching methods and innovations in the evaluation techniques is also ascertained through experimental research.

STEPS IN EXPERIMENTAL RESEARCH

The steps of the experimental method are not different from those of a scientific method. For the sake of clarification, the major steps may be described as under:

Surveying the Literature Relating to the Problem

For a worthwhile research based on experimentation, the researcher like in any other type of research, needs to acquire up-to-date information relating to his problem. The sources and the procedures which are helpful to acquaint the researcher with adequate related literature have been discussed earlier in Chapter Four.

Selecting and Defining the Problem

Experimental research starts with the selection of the problem which is amenable to experimentation. It needs a rigorous logical analysis and definition of the problem in precise terms. The variables to be studied should be defined in operational terms clearly and unambiguously. It helps the researcher to convert the problem precisely into a hypothesis that can be verified or refuted by the experimental data.

Stating of Hypotheses

The stating of problem hypotheses is one of the distinguishing characteristics of the experimental method. Hypotheses are the heart of experimental research. They suggest that an antecedent condition or phenomenon (independent variable) is related to the occurrence of another condition, phenomenon, event, or effect (dependent veriable). To test a hypothesis, the researcher attempts to control all the conditions except the independent variable which he manipulates. Then he observes the effect on the dependent variable presumably because of the exposure to the independent variable. The researcher, therefore, should not only be concerned primarly with experimental plans and statistical procedures, but should give sufficient attention to the formulation of hypotheses. The experimental plans and statistical procedures merely help him in the testing of hypotheses and contribute little in the development of theories or advancement of knowledge. The hypotheses developed or derived from existing theories, however, contribute to the development of new theories and knowledge.

Constructing the Experimental Plan

Experimental plan refers to the conceptual framework within which the experiment is conducted. According to Van Dalen (1973, p. 260), it represents all elements, conditions or phenomena, and relations of consequences so as to:

(*i*) identify all non-experimental variables that might contaminate the experiment and determine how to control them;
(*ii*) select a research design;
(*iii*) select a sample of subjects to represent a given population, assign subjects to groups, and assign experimental treatments to groups;
(*iv*) select or construct and validate instruments to measure the outcomes of the experiment;
(*v*) outline procedures for collecting data and possibly conduct a pilot or 'trial run' test to perfect the instruments or design ; and
(*vi*) state the statistical or null hypothesis.

The identification of non-experimental or extraneous variables and the procedures for controlling them have already been discussed earlier in this chapter. In order to select a suitable research design for conducting the experiment and assign subjects to different experimental treatments to measure the outcomes of experiment, the researcher must be well acquainted with different types of experimental designs.

EXPERIMENTAL DESIGNS

An experimental design is to the researcher what a blueprint is to an architect. It provides the researcher an opportunity for the comparisons required by the hypotheses of the experiment and enables him to make a meaningful interpretation of the results of the study with the help of statistical analysis of the data. There are three important criteria which the researcher must keep in mind while selecting an experimental design for conducting his experiment.

Appropriateness

The first and the most important criterion is that the design should be appropriate for testing the hypotheses of the study. If the design is not appropriate, the results of the study will not be worthy of serious consideration. The important criterion of a well designed experiment is not complexity or simplicity but appropriateness. The researcher, therefore, should select the design that will do the job it is supposed to do and is able to arrange objectively the experimental conditions to meet the requirements of his study.

Adequacy of Control

The second criterion is that the design must provide 'adequate control' so that the effects of the independent variable on the dependent variable can be measured. The adequate control of extraneous variables helps the researcher to get dependable answers to the questions raised by the hypotheses of the study. As mentioned earlier, 'randomization' is the only best method of controlling all possible extraneous variables. Ary *et al.* (1972, p. 231) advise to select a design that utilizes randomization in as many aspects as possible. They are of the opinion that:

> If it is not possible to select subjects at random, then try to assign subjects to groups randomly. If neither of these is feasible, then at least an attempt should be made to assign experimental treatments to the groups at random.

The design with adequate control suggests what observations are to be made by the researcher, how to make them, what statistical tests are feasible, how to analyse the obtained experimental data, and what results may be drawn from the statistical analysis.

Validity

The third criterion that the researcher must give attention to is the 'validity' of the design. It is essential for the purpose of testing the particular hypothesis of the study. Campbell and Stanley (1963) have suggested that there are two general types of validity: internal and external.

Internal validity. One of the major objectives of the researcher in experimentation is to determine whether the variables that have been identified actually have a systematic effect on the dependent variable and whether the observed results were not affected by the extraneous or situational variables. The extent to which this aim is attained is a measure of internal validity of experiment. This validity is basically a problem of control. The extraneous variables that affect the control of a design contribute to its internal validity.

Campbell and Stanley (1963) have pointed out that there are eight extraneous variables which affect significantly the internal validity of a research design. They are of the opinion that these variables must be controlled or else they might lead to alternative interpretations of the results of the experimental study.

(*i*) *History.* The researcher should try to control the specific events, other than the experimental treatment, that may occur between the first and the second measurements of the subjects to affect the dependent variable.

(*ii*) *Maturation.* The time period that elapses during the experimentation may produce certain changes in the subjects. For example, the subjects may perform differently on the dependent variable on different occasions as a result of biological or psychological processes like fatigue, age, interest or motivation. Therefore, the effects on the dependent variable as a result of the change in subjects due to passage of time could mistakenly be attributed to the experimental variable.

(*iii*) *Pre-testing.* The exposure of the subjects to the pre-test may serve as the learning experience and therefore it may affect their post-test performance.

(*iv*) *Measuring instruments.* Different measuring instruments, scorers, raters, interviewers or the observers used at the pre-and post-testing stages may also account for the observed differences in the scores or measures of the dependent variable.

(*v*) *Statistical regression.* The groups chosen on the basis of extreme scores may cause statistical regression effect. It refers to the tendency for extreme scores to regress or move towards the common mean on subsequent measures. Such a tendency may operate to produce an effect that could be interpreted as an effect due to experimental treatment. Suppose that the subjects who do exceptionally poorly on a proficiency test in Hindi are selected to receive an experimental treatment in the form of special coaching in Hindi. The mean score of the group will tend to move towards the mean of the parent population on the second test whether or not an experimental treatment is applied. Similarly, the initial mean of the subjects who do exceptionally well on the proficiency test in Hindi would tend to go down towards the population mean on the second testing after receiving the experimental treatment.

(*vi*) *Differential selection of subjects.* The groups may differ significantly on some important variables related to the dependent variable even before the application of the experimental treatment. For example, if the subjects in the experimental group in an experiment of retention are more intelligent than the subjects in the control group, the former may perform better on the dependent variable (retention measure) even if this group did not receive an experimental treatment.

(*vii*) *Experimental mortality.* The differential loss of subjects from the comparison groups may affect the findings of the study. If, for examples, some subjects in the experimental group who receive the lowest scores on the pre-test drop out after taking the test, this group will show higher mean on the post-test than the control group, not because of the experimental treatment but because the low scoring subjects are not present

(*viii*) *Interaction of selection and maturation, selection and history, etc.* When the two comparison groups have the same scores on the pre-test, some other differences due to interaction between the variables such as intelligence, motivation, interest, age, etc., rather than experimental variable may cause one of the groups to get higher post-test scores. Such interaction occur when subjects are selected into groups on the basis of factors extraneous to the purpose of the experiment.

External validity. The second important objective of the researcher is to determine whether the systematic relationships that have been identified, isolated and measured can be generalized outside the experimental setting. The extent to which this objective is attained is a measure of the external validity of the experiment. This validity is concerned with the generalizability or representativeness of the experimental findings, that is, to what subject populations, settings, experimental variables and measurement variables can the results of the experiment be generalized.

Bracht and Glass (1968) have classified external validity into two types: (*i*) population validity and (*ii*) ecological validity.

(*i*) *Population validity*. Population validity is concerned with the identification of the population to which the results of an experiment can be generalized. For example, a researcher has compared the effectiveness of programmed instruction and conventional method in teaching social studies to a sample of seventh grade students in Himachal Pradesh and found that programmed instruction is more effective. From this result, the researcher would like to conclude that programmed instruction is superior for other groups of seventh grade students in Himachal Pradesh. In order to make valid generalizations from the experimental results to larger populations, the researcher must correctly identify populations to which the results can be generalized. For this the researcher has to make distinction between the experimentally accessible population and the target population. The former refers to the population of subjects that is within the reach of the researcher for his study. The latter is the total group of subjects to whom he wants to apply the conclusions from the results of his experiment. In the previous example, all the seventh grade students studying in the middle and high schools of Simla District of Himachal Pradesh may be the experimentally accessible population. The target population would be all seventh grade students in Himachal Pradesh. The generalizations of the study would be done in two phases: (1) from the experimental sample of seventh grade students to the experimentally accessible population (seventh grade students of all middle and high schools of Simla District), and (2) from the accessible population to the target population (seventh grade students of all middle and high schools of Himachal Pradesh).

If the researcher has strictly followed the principle of randomization in selecting the experimental sample from the experimentally accessible population (seventh grade students in the middle and high schools of Simla District), then he can generalize the findings to this population with no difficulty. For this the researcher has to specify the accessible population by listing and numbering its every member and then select the sample through the use of random numbers or any other random procedure

Generalizations from the accessible population to the target population is somewhat risky and cannot be made with the same degree of confidence as the former type. For such generalization, the researcher must have a thorough knowledge of the characteristics of the accessible and the target populations. If the characteristics of both the populations are similar, the researcher can generalize the results with more confidence. In the previous example, if the researcher has randomly selected his experimental sample from all seventh grade students of Simla District, then the accessible population would be more like the target population (all seventh grade students of Himachal Pradesh) and the findings could be generalized to the target population with much more confidence. When the researcher, therefore, attempts to generalize from accessible population to the target population, it is important for him to know that one is similar to the other with respect to certain significant and relevant characteristics.

The interaction as the result of "selection by treatment" also contributes to population validity. When a researcher selects two experimentally accessible populations and these are not representative of the same target population, the similar studies on two accessible populations can lead to entirely different results. In other words, an interaction that may occur between the treatment and the characteristics of one experimental sample selected from one accessible population would not occur in another experimental sample selected from the second accessible population with different characteristics. Therefore, it would not be possible for the researcher to generalize the findings from one sample to another.

(*ii*) *Ecological validity*. In addition to population external validity, the researcher should also be concerned with ecological external validity. It is concerned with generalizing experimental effects to other environmental conditions. Ary *et al.* (1972, p. 235) suggest that:

> To have ecological validity, a design must provide assurance that the experimental effect is independent of the particular experimental environment.

The authors are of the opinion that the researcher must give increased attention to the following factors for achieving ecological validity:

1. The researcher must furnish a complete description of the operations and the experimental setting involved in the experimental study. It helps a reader to judge to what extent the results can be generalized to other situations. For example, a researcher has studied the effectiveness of programmed instruction in comparison to lecture method in the teaching of social studies to fifth grade students under certain experimental conditions. The findings would not be applicable to other subject areas of the same grade, unless the study is extended to a variety of teaching situations. One may ask if the findings of the study would be the same if a school in rural setting were used instead of an urban school in a replication of the study and so on.

2. The researcher must give due consideration to the experimental

arrangements while generalizing the findings of his study. There may be reactive effect due to the experimental procedures. The presence of observers, experimental equipment and knowledge of participation of the subjects in an experiment makes the subjects aware of the fact that they are receiving experimental treatments and, therefore, they may change their normal behaviour. Such a change is called 'Hawthorne effect'. If subjects change the behaviour that is being measured, the researcher cannot claim that the effect of treatment variable for the same population will be the same for subjects who are exposed to the treatment variable in non-experimental situations.

3. Another concern in ecological validity is the question of the representativeness of the independent or experimental variables and dependent or criterion variables. Variable representativeness influences the generalizability of the findings of an experimental study. For example, when a researcher speaks of aggression, he may ask himself "what kind of aggression does he mean?"

4. Certain interaction effects may also influence the generalizability of experimental findings. When two or more treatments are administered to the same group within the same or different studies, it is difficult to find the cause of the experimental results or to generalize the results to the experimental settings in which only one treatment is present. In some situations, a pre-test may increase or decrease the sensitivity or responsiveness of experimental subjects to the independent or experimental variable and, therefore, findings obtained for this pre-tested population cannot be generalized for the unpretested population from which the experimental subjects were selected.

5. The tools or instruments used for the measurement of dependent variable also contribute to the ecological validity. For instance, if a researcher uses an essay type test to measure academic achievement (dependent variable), can one say that the same effect would be observed if an objective type test were used as the measure of dependent variable?

The researcher must select a design that is strong both in internal and external validity. However, in certain educational experiments, one type of validity can be obtained at the cost of another one. In such cases, the researcher should attempt to reach a compromise between two types of validities, i.e. he should choose a design, within practical limits, that provides sufficient control to make results interpretable and generalize findings to the intended settings.

Some Terms and Symbols

Certain terms and symbols are commonly used in the discussion of various types of experimental designs. It is necessary for the researcher to become familiar with them.

1. *Experimenter*. The person who conducts the experiment or manipulates the experimental conditions is called the experimenter. It is denoted by E.

2. *Subject*. The living organism that is studied is called the subject or the respondent. It is denoted by S and the plural is Ss.

3. *Independent variable*. The variable which is manipulated by the experimenter is called independent variable. It is often called the treatment, experiment, or antecedent variable, and is represented by the symbol X. In educational research an independent variable may be a particular teaching method, a type of teaching aid or material, a reward or punishment, or period exposure to a particular condition.

4. *Dependent variable*. The dependent variable is the condition or characteristic that appears, disappears, or changes as the experimenter introduces, removes, or changes independent variable. It is also called criterion or predicted variable and is symbolized by the letter Y. The dependent variable may be a test score, the number of mistakes or errors, measured speed in performing a task. The symbol Y_1 is used to represent the dependent variable before the manipulation of the independent variable X and Y_2 represents the dependent variable after the manipulation of the independent variable.

5. *Control group*. The group that does not receive any experimental treatment is called the control group. It is the group that is not exposed to some independent variable X, or is exposed to another independent variable X for comparison purposes.

6. *Experimental group*. The group that is given the independent-variable treatment or is exposed to some independent variable X is called the experimental group.

7. *Pre-test*. The test that is administered to the subjects before the independent variable X is called the pre-test. It is represented by the symbol T_1.

8. *Post-test*. The test that is administered to the subjects after the independent variable be is applied is called the port-test. It is represented by the symbol T_2.

TYPES OF EXPERIMENTAL DESIGNS

There are various types of experimental designs. They vary in complexity and adequacy. The selection of a particular design depends upon such factors as the nature and purpose of the experiment, the type of the variables to be manipulated, the nature of the data, the facilities or the conditions for carrying out the experiment, and the competence of the experimenter. Although the designs can be combined into various ways, they are broadly classified as under:

1. Pre-experimental designs
2. True experimental designs
3. Factorial designs
4. Quasi-experimental designs
5. Time-series designs

The designs resemble one another from the point of view of purpose and their adherence to the principles of experimentation. They differ in the degree of accuracy with which they attack the problem or meet the essential critreia of control, manipulation, observation, and replication. No design solves all the problems. The nature of the problem determines which type of design is most appropriate and applicable and how the design should be used to meet the requirements of the experiment.

Pre-experimental Designs

Pre-experimental designs provide little or no control of extraneous or situation variables. They are, however, still being used in the study of educational problems.

There are two types of pre-experimental designs.

Design 1: One Group Pre-test Post-test Design

When an experimenter uses this design, he measures dependent variable, before the independent variable *X* is applied or withdrawn and then takes its measurement again afterwards. The difference in the measurements of dependent variable, if any, is computed and is taken as the amount of change as a result of the application or withdrawing of independent or treatment variable.

Illustration

Suppose a teacher wants to evaluate the effectiveness of programmed instruction in teaching general science to sixth grade students. At the beginning of the session he administers a criterion test to a selected group of sixth grade students that seems to be the good measure of the achievement of the objectives in general science for sixth grade and gets the measure T_1. Then the group is administered the programmed text on general science, and at the end of the session the teacher administers the criterion test a second time to get the measure T_2. The means of T_1 and T_2 are compared to ascertain what difference, if any, the exposure to X (teaching through programmed instruction) has made. An appropriate statistical technique will be used to ascertain whether the difference is statistically significant.

PARADIGM FOR THE DESIGN 1: One Group Pre-test Post Design

Pre-test	*Independent variable*	*Post-test*
T_1 Mean of the criterion test	X Teaching through programmed instruction	T_2 Mean of the criterion test

Limitations

Since the design involves only one group and one teacher, it seems to control intersubject differences and extraneous variables. The control, however, is superficial and does not check the threats to internal validity because of the following reasons:

1. This design does not use any control group and, therefore, the experimenter cannot assume that the difference between the pre-test mean T_1 and the post-test mean T_2 was brought about by the experimental treatment or by some extraneous variables.

2. History and maturation are two major extraneous variables that are not controlled in this design. History refers to the specific events that can occur between the pre-test and the post-test other than the exposure of subjects to the experimental treatment. In the previous illustration, for example, the increased emphasis on general science in the school, or the teaching of the group by an effective general science teacher during experimentation could increase student achievement in this area. Maturation variable refers to changes in the subjects themselves that occur with the passage of time. For example, the subjects grow mentally as well as physically and their learning experiences between the pre-test and post-test could affect the dependent variable. Ary *et al.* (1972, p. 240) are of the opinion that history and maturation become potent sources of extraneous variance when the time interval between the pre-test and post-test is long.

3. This design does not provide any procedure for evaluating the effect of post-test itself. There is practice effect when the subjects take a test a second time or even take a parallel form of the test. That is, subjects perform better at the post-test stage even without any teaching.

4. There is a problem of reactivity in the design due to a reaction between the subject and pre-test measure. It is this reaction rather than the treatment variable that produces the change in the post-test measures. For example, the novel or controversial content of a pre-

test may motivate the subjects to react in a particular manner and it is this reaction that brings about the observed change in subjects at the post-test stage.

Design 2: Two Groups, Static Design

To overcome the limitations of Design 1, 'two groups, static design' utilizes two groups, only one of which is exposed to the experimental treatment. The other group which is not exposed to any experimental treatment acts as the control group and this permits the comparison that is required by a scientific investigation. The experimenter assumes the two groups to be equivalent in all relevant aspects at the start of the experimentation. There is no pre-test and the dependent variable measures for the two groups are then compared (post-test), to determine the effect of independent or treatment variable X. If the experimental group performs better on the post-test (T_2), the experimenter s more confident that the independent variable is more responsible for the change in the dependent variable.

Illustration

Suppose a teacher wants to evaluate the effectiveness of a new method teaching in teaching arithmetic to fifth grade students. He will frame two equivalent groups of 20 each from the same class of 40 students. One group will be taught by the new method and the other one by the conventional method. At the end of the experiment, an achievement test in arithmetic is administered to both the groups and a comparison is made of the achievement of students of the two groups. The means of the post-test T_2 will be computed. An appropriate statistical test will be applied to ascertain whether the observed difference is statistically significant.

PARADIGM FOR DESIGN 2: **Two Groups Static Design**

Group	*Independent Variable*	*Post-test*
Experimental	Teaching through new method	T_2
Control	Teaching through conventional method	T_2

Limitations

Since neither randomization nor matching is used to assign subjects to the experimental and control groups, the experimenter cannot assume that the groups are equivalent with respect to relevant extraneous variables before they are exposed to the experimental treatment. This design, therefore, is also considered to be lacking in the necessary control. Any differences between the post-test measures T_2 of the experimental and control groups may come about through the biased selection or inclusion of the subjects forming the groups, or they may differ on certain relevant variables without the effect of the experimental variable. Moreover, experimental mortality may also produce differences in the groups due to differential dropout of subjects from the groups. Therefore, the two groups once equivalent may differ later because of selective dropout of subjects.

True Experimental Designs

True experimental designs are mostly used for experimental research in education because they seek to control the main effects of history, maturation, testing, measuring instruments, statistical regression selection, and mortality.

True experimental designs can be classified into four types.

Design 3: Two Groups, Randomized Subjects, Post-test-only Design

This design is one of the simplest and powerful experimental designs. The available subjects are assigned to two groups through randomization which controls for all possible relevant extraneous variables. No pre-test is used and the random assignment of subjects assures that any initial differences between the groups are attributable only to chance. The two random samples from designed population are obtained in two ways: (1) The subjects may be drawn individually at random and assigned alternatively to the groups; or (2) two different random samples may be selected first and the groups are assigned randomly to the experimental or control condition by the flip of a coin. Only the experimental group is exposed to the experimental treatment. At the end of the experiment, subjects of both the groups are measured on the dependent variable T_2. The means of the two groups are compared with the help of appropriate statistical test of significance.

Illustration

Suppose an experimenter wants to ascertain whether a new teaching method will increase reading speed of third grade students. He prepares a list of all the elementary schools of a particular city and assign numbers to all the third grade students. With the help of random number table he may draw a desired sample of 100 students Then two random samples of 50 subjects each may be selected from the selected sample in two way : the researcher may select subjects individually at random and assign them alternately to the groups, or he may first draw two random samples and then assign groups to the experimental or control condition by tossing a coin.

After assigning the subjects to two groups, the experimental group is taught through the new method, and the control group through the conventional method, for a period of time. In all other respects, the researcher will treat the groups alike. After the desired period of time, the subjects of both the groups will be administered a reading test. The mean scores of the two groups are compared to determine the effectiveness of new teaching method by using an appropriate statistical test. If the obtained means of the groups are significantly different, the experimenter can be reasonably confident that the use of new teaching method was responsible for the observed difference.

PARADIGM FOR THE DESIGN 3: Two Groups, Randomized Subjects, Post-test-Only Design

Randomly assigned group	*Independent variable*	*Post-test*
	X	
Experimental	Teaching through new method	T_2
Control	Teaching through conventional method	T_2

Advantages

1. The main advantage of this design is randomization, which assures statistical equivalence of the groups prior to the introduction of the experimental treatment.

2. Since no pre-test is used, this design controls for the main effects of history, maturation, and pre-testing. Moreover, there can be no interaction effect of pre-test and independent or experimental variable. Hence, this design is especially recommended for the experiments in which pre-test sensitization is likely to occur.

3. This design is useful in the experimental studies, especially at kindergarten or primary stages, in which a pre-test is either not available or not appropriate.

4. If necessary, this design can be extended to include more than two groups.

Limitations

In spite of all the advantages mentioned above, this design suffers from the following limitations:

1. The use of this design seriously restricts the external validity of the experiment. The experiment can partially overcome this limitation by replicating the experiment with different groups.

2. There are some situations in which it is not possible for the

experimenter to select subjects at random from the population of interest. In such cases, Ary *et al* (1972, p. 243) suggest that the experimenter must begin with available subjects and assign them randomly to the group.

Design 4: Two Groups, Randomized Matched Subjects, Post-test-Only Design

This design instead of using random assignment of subjects to experimental and control groups, uses a technique of matching. The variables selected for matching must have a significant correlation with the dependent variable and can be measured conveniently. The pre-test scores on the dependent variable or the criterion, if available, can be used very effectively for the matching procedure. The subjects from the desired population are paired so that their scores on the matching variable are as close together as possible. One subject of each pair is randomly assigned to one group and the other to the second group. A coin is tossed to designate the groups as experimental and control groups. The experimental group is given the experimental treatment. After the treatment is over, subjects of both the groups are measured on the dependent variable T_2. The significance of the difference between the two means is ascertained with the help of appropriate statistical technique.

Illustration

Suppose an experimenter is interested in studying the effect of interim tests on the achievement of seventh grade students in general science. He randomly selects 100 students, from the population of seventh grade students studying in the high schools of some city. Intelligence and previous knowledge of general science will be the relevant matching variables that have a significant correlation with the achievement in general science (dependent variable). The experimenter will select the pairs of students from the desired population in such a way that the scores on pre-test (achievement in general science) and intelligence test of the students of each pair are as close together as possible. Then one student of each pair is randomly assigned to a group which will receive instruction with interim tests (*i.e.* the group which will be administered interim tests at the end of each unit of the selected content in general science) and the other to the second group which will receive instruction without interim tests. At the end of the experiment, students of both the groups will be administered an achievement test on the selected content in general science. The means of the achievement scores for both the groups will be computed to test the significance of the observed difference between them. If the difference comes out to be significant, the experimenter will conclude with confidence that the observed difference in performance is due to the effect of interim tests administered during the classroom instruction.

PARADIGM FOR THE DESIGN 4: Two Groups, Randomized Matched Subjects Post-test-Only Design

Randomly assigned group after matching	*Independent variable*	*Post Test*
Experimental	Instruction with interim tests	T_2
Control	Instruction without interim tests	T_2

Advantages

1. This Design is most useful where Design 3 is not applicable and where small groups are to be used.

2. The random assignment of subjects to the experimental and control groups after matching adds to the strength of this design.

3. Design 3 is based on the random selection of subjects to obtain identical groups. With small groups it does not provide any assurance that the groups are really comparable because of the pre-existing inter-subject differences on variables highly related to the dependent variable that the experimental treatment is applied to affect. Design 4, however, controls all such pre-existing inter-subject differences.

Limitations

1. This design is subject to the limitations that have been mentioned earlier while discussing matching as a means of control.

2. In some situations it is not possible to locate a match of one or more potential subjects. A bias gets introduced in the sample as a result of exclusion of such cases from it.

Design 5: Randomized Groups, Pre-test-Post-test Design

This design is also called as 'Randomized control-group Pre-test-Post-test Design'. In this design, subjects are assigned to the experimental and control groups by random procedures and administered a pre-test T_1 as a measure of the dependent variable Y. The experimenter introduces the treatment only to the experimental group for a specified period of time. At the end of the experiment, the experimental and control groups are administered the post-test T as the measure of dependent variable. The difference between means of T_1 and T_2 is found for each group and these mean difference scores are compared with the help of an appropriate statistical test in order to ascertain whether the experimental treatment produced a significant effect than the control condition.

Illustration

Suppose an experimenter wants to study the effectiveness of 'Structural Approach' in teaching English to sixth grade students, first he will select subjects from a population of sixth grade students by random methods and then randomly assign subjects to experimental and control groups. A pre-test measuring dependent variable (performance on an achievement test) will be administered on the groups to obtain T_{1E} scores for the subjects of the experimental group and T_{1C} scores for the control group subjects. Keeping all the conditions same for the two groups, experimental group will be taught through the structural approach and the control group by conventional method for a stipulated period of time. At the end of the instruction, the experimenter will test the subjects of the groups on the dependent variable to obtain T_{2E} scores for the experimental group subjects and the T_{2C} scores for the control group subjects. The difference between the T_1 and T_2 scores for each subject and the mean of these differences for each group, D_E and D_C, will be determined. To ascertain whether the performance in English of the experimental group is sufficiently great as a result of teaching through structural approach, appropriate statistical test will be applied to test the significance of the difference $(D_E—D_C)$.

PARADIGM FOR THE DESIGN 5: Randomized Groups, Pre-test Post-test Design

Randomly assigned	*Pre-test*	*Independent variable*	*Post-test*
Experimental group	T_{1E}	Teaching through structural approach	T_{2E}
Control group	T_{1C}	Teaching through traditional method	T_{2C}

D_E=mean of the difference between experimental subject's pre-test and post-tes scores.

D_C=mean of the difference between the control subject's pre-test and post-tes scores.

Compare D_E and D_C to ascertain effect of teaching through structural method.

through structural approach). For example, even though the experimental group has a significantly higher mean on post-test (T_{2E}) than the first control group (T_{2C_1}), the experimenter cannot be confident that this difference is due to its teaching through structural approach. However, if the post-test mean of the second control group (T_{2C_2}) is also significantly higher than that of the first control group (T_{2C_1}), then the experimenter can assume that teaching experimental group through structural approach has produced the difference rather than the interaction due to post-test and experimental treatment, because second control group (C_2) was not administered any post-test.

PARADIGM FOR THE DESIGN 6: **The Randomized Solomon Three-Group Design**

Randomly assigned	*Pre-test*	*Independent variable*	*Post-test*
Experimental group (E)	T_{1E}	Teaching through structural approach	T_{2E}
Control group (C_1)	T_{1C_1}	Teaching through conventional method	T_{2C_1}
Control group (C_2)	No pre-test	Teaching through structural approach	T_{2C_2}

Advantages

1. The main advantage of this design, like Design 5, is the initial randomization which assures equivalence between the groups prior to experimentation. Therefore, it controls most of extraneous variables, such as history, maturation, pre-testing, differential selection of subjects, statistical regression and mortality.

2. The addition of second control group provides control over the interactive effect of pre-testing and the experimental treatment.

Limitations

This design does not have control over any possible contemporary effect that may occur between: (*i*) T_{1E} and T_{2E}; (*ii*) T_{1C_1} and T_{2C_1}.

Design 7: The Randomized Solomon Four-Group Design

This design overcomes the external validity weakness which exists in Design 5, and provides still more rigorous control by extending Design 6 to include one more control group.

In this design the subjects are assigned at random to the four groups. The experimental group and one of the control groups are administered a pre-test. The other two control groups are not pre-tested. The design enables the experimenter to control and measure both the main and interaction effects of testing. Moreover, the main effects of a composite of maturation and history are controlled in this design.

Illustration

In the example illustrated in Design 5, the experimenter will frame four groups instead of two groups. He will designate one group as the experimental group (E) and the other three as control groups. The pre-test T_1 will be administered to the experimental group and one of the control groups to obtain the measures T_{E_1} and T_{C_1} respectively. The other two control groups will not be pre-tested.

The experimental group (E) and the second control group will be taught through structural approach. The first and the third control groups get instruction through conventional method for a stimulated period of time. At the end of instruction, the subjects of the three groups will be measured on the dependent variable to obtain T_{2E} scores for the experimental group and T_2C_1, T_2C_2 and T_2C_3 scores for the control groups.

The effectiveness of teaching through structural approach will be ascertained by the comparison of mean scores of T_2 *i.e.* T_{2E}, T_2C_1, T_2C_2 and T_{2C_3}. For this appropriate statistical test will be used by the experimenter.

The experimenter can make several comparisons to determine the effect of teaching through structural approach. For example, if the post-test mean (T_{2E}) of the experimental group (E) is significantly greater than the mean (T_2C_1) of the first control group (C_1) and if the post-test mean (T_2C_2) of the second control group is significantly greater than that of the post-test mean (T_2C_3) of the third control group, the experimenter can conclude that the instruction through structural approach is more effective than the conventional teaching. Moreover, the influence of the instruction through structural approach on a pre-tested group can be determined by comparing the post-test measures T_{2E} and T_{2C_1} of the experimental and first control group respectively or pre-post test changes of these two (E and C_1) groups. The effect of the experimental conditions (instruction through structural approach) on an unpretested group can also be determined by comparing the second and third control groups (C_2 and C_3). If the experimenter finds that the average differences between post-test scores $T_{2E}—T_{2C_1}$ and $T_{2C_2}—T_{2C_3}$ are about the same, then he may infer that the experiment has a comparable effect on pre-tested and unpre-tested groups.

PARADIGM FOR DESIGN 7: The Randomized Solomon Four-GroupDesign

Randomly assigned	*Pre-test*	*Independent variable*	*Post test*
Experimental group (E)	T_{1E}	Teaching through structural approach	T_{2E}
Control group (C_1)	T_{1C_1}	Teaching through conventional method	T_{2C_1}
Control group (C_2)	No pre-test	Teaching through structural approach	T_{2C_2}
Control group (C_3)	No pre-test	Teaching through conventional method	T_{2C_3}

Advantages

1. In addition to the advantages of Design 6, the Design 7 provides control over any possible contemporary effects that may occur between pre-testing and post-testing.

2. This design actually involves conducting the experiment twice, once with pre-tests and once without pre-tests. If the results of these two experiments are in agreement, the experimenter can have much greater confidence in his findings.

Limitations

1. The experiment involving Design 7 is difficult to carry out in practical situations. It involves more time and effort to conduct two experiments simultaneously and there is the problem of locating the increased number of identical subjects that would be required in the experiment.

2. Since this design involves four sets of measures for four groups and the experimenter has to make comparisons between the experimental and first control group (E and C_1) and between second and third control groups (C_2 and C_3), there is no single elementary statistical procedure that would make use of the six available measures simultaneously. In the light of this difficulty, this design is generally recommended for a more advanced level of research.

Factorial Designs

The discussion about the various designs thus far has been confined to classical single variable designs which require that an experimenter manipulate one independent variable to produce an effect on the dependent variable. Because human nature is complex, an experimenter in educational situations cannot always fulfil these requirements. The independent variable alone may not produce the same effect as it might in interaction with another independent variable. The findings, therefore, from a one-variable design may be meaningless. For example, the effectiveness of a new teaching technique depends upon a number of variables such as the intelligence level of students, the type of the teacher teaching the group, and so on. In such an experiment, a classical one-variable design would not reveal any information about the interactive effect of the method of teaching and intelligence level or any other variable.

R.A. Fisher overcame this difficulty when he developed factorial designs and the statistical techniques like analysis of variance and covariance for the study of complex interactions.

A factorial design enables the experimenter to evaluate or manipulate two or more variables simultaneously in order to study the effects of number of independent factors singly as well as the effects due to interactions with one another. Factorial designs vary according to the degree of complexity depending upon the nature and purpose of the experiment. They include two or more independent variables, and each one is manipulated in two or more ways to assess both their separate (main) and their combined (interaction) effects.

Design 8: Simple Factorial Design of 2 by 2 (2 × 2)

The simplest factorial design is 2 by 2 (2 × 2). In this design there are two independent variables and each of the independent variables has two values.

The first independent variable, which is manipulated and has two values, is called the experimental variable. The second independent variable, which is divided into levels, may be called the control variable.

Illustration

Suppose an experimenter is interested in comparing the effectiveness of programmed instruction and lecture method—methods *P* and *L*—on the achievement in social studies of sixth grade students. From the survey of related literature, he comes to believe that there may be a differential effect of these methods on different levels of intelligence of the students. On the basis of IQ, the experimenter divides the population into two groups: one group of high intelligence level and second group of low intelligence level. He randomly selects 40 subjects from the group with high intelligence level and assigns 20 subjects to method *P* and 20 subjects to method *L*. He adopts the same procedure for the low intelligence group. He also randomly assigns teachers to each of the group.

In this experiment, there are two experimental treatments, that is, teaching through programmed instruction and teaching through lecture method. Moreover, there are two levels of intelligence. Therefore, there are four groups of subjects within each of the two levels of intelligence and are randomly assigned to the two treatments. The factorial design is shown in the Figure 12.1.

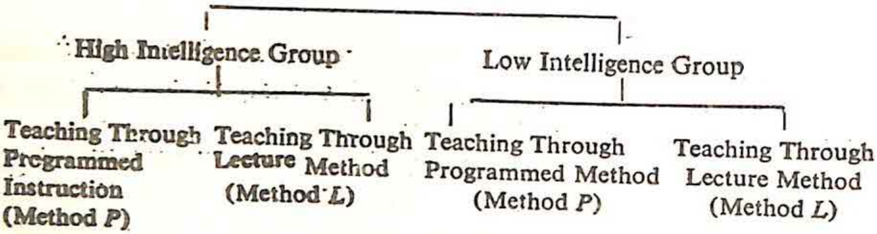

High Intelligence Group
Low Intelligence Group
Teaching Through Programmed Instruction (Method *P*)
Teaching Through Lecture Method (Method *L*)
Teaching Through Programmed Method (Method *P*)
Teaching Through Lecture Method (Method *L*)

FIG. 12.1. Sketch Illustrating the Treatments at Two Levels of General Intelligence.

Both the groups are given instruction for the stipulated period. At the end of the instruction, the groups are administered an achievement test in social studies and the data are arranged in the respective categories to compute mean achievement score for each of the four groups. The mean scores are recorded in the appropriate square (cells) in the Table 12.1. Since one of the objectives of the experimenter is to compare various combinations of these groups, he also obtains the mean scores for each row and for each column. The marginal column means are for the two methods and the marginal row means are for the two levels of intelligence.

TABLE 12.1: Example of a Factorial Design

	Method (X_1)		
Intelligence level (X_2)	*Method L*	*Method P*	*Mean*
High	76.00	74.00	75.0
Low	61.00	65.00	63.0
Mean	68.5	69.5	—

Main Effects

From the Table 12.1, the experimenter can first determine the main effects for the two independent variables. The treatment mean scores without regard to intelligence levels indicate the main effect for treatments given in the form of teaching methods (X_1). The mean of the high and low intelligence groups taught through lecture method

(Method *L*) is 68.5 and that of the groups taught through programmed instruction (Method *P*) is 69.5, and the difference between means is 1.0. Therefore, the experimenter could conclude that teaching through programmed instruction is more effective than the lecture method.

In order to determine the main effect of intelligence level (X_2) on the achievement scores, the mean 75.0 for the high intelligence level group is compared with mean 63.0 for the low level intelligence group. This main effect for intelligence levels does not take into account any differential effect due to teaching methods.

Since the mean of high intelligence level group is higher in comparison to low intelligence level group, it may be inferred that, regardless of the teaching method, the former group has performed better than the latter group.

Interaction Effect

In addition to the main effects, the experimenter is able to assess the interaction between the intelligence level and teaching method. It enables him to assess the differential effects of one of them at different levels of the other. This interaction effect is shown in the Figure 12.2.

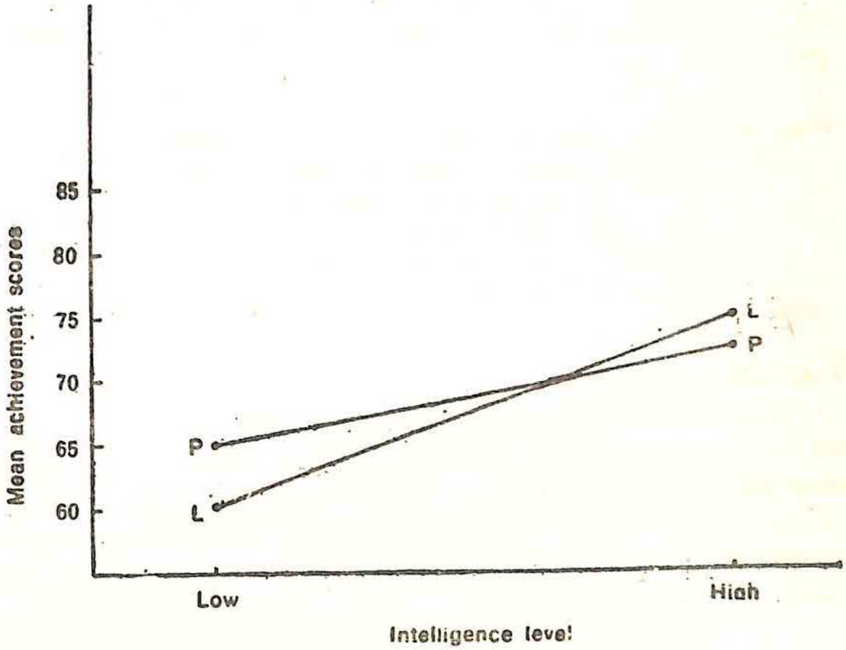


FIG. 12.2. Illustration of Interaction between Method and IQ Level.

If there is an interaction, the effect of the teaching method on achievement will differ for the two intelligence levels. In such a situation, the lines shown in the figure intersect. If there is no interaction, the effect of the teaching method on the achievement of the students will be the same for both levels of intelligence. The lines in such a case run more or less parallel. From an examination of the Figure 12.2 and the Table 12.1, it may be inferred that teaching through the programmed instruction (Method *P*) is more effective than lecture method (Method *L*) for the low intelligence group, and the reverse is true for high intelligence group.

Advantages

1. The differences in the effect of different levels or categories of more than one variable can be studied with factorial designs simultaneously. An experimenter, therefore, can accomplish in one experiment what otherwise might require two or more separate experiments.
2. While studying the significance of the differences in the dependent variable under the effect of the levels of any of the factors the groups become alike with respect to the different levels of the other factors and thus the groups get controlled as far as the levels of the other factors are concerned.
3. Besides studying the significance of the differences in the levels of the factors, the factorial designs provide an opportunity to study interactions between the factors. In the words of Edwards (1971, p. 226).

> If the interactions involving a given factor are not significant, then we obviously have a broader basis for generalizing about the main effect of the factor, because it has been tested in conjunction with variations of other factors rather than holding the other factors constant at arbitrary levels. If, on the other hand, we have a significant two-factor interaction, examination of the nature of the interaction may provide us with additional insight as to how each factor operates.

Limitations

A factorial design may include any number of independentv variables with any number of levels of each. However, when the experimenter manipulates or controls too many factors simultaneously, the experiment and the statistical analysis of the data sometimes become unmanageable. Moreover, the combinations of too many variables also become artificial.

Quasi-Experimental Designs

The true experimental design, as discussed earlier, provide full experimental control through the use of randomization procedures.

There are many experimental situations in which it is not possible for the experimenter to assign subjects randomly to groups or exercise full control over the scheduling of experimental conditions. In such situations, he uses quasi-experimental designs, that provide as much control as possible under the existing conditions. If an experimenter uses a quasi-experimental design, it is necessary for him to know which of the variables his design may fail to control. He must also be aware of the sources that represent threats to both internal and external validity and consider them while interpreting the results of the experiment.

Some of the important quasi-experimental designs are dicussed as under:

Design 9: Non-randomized Control Group, Pretest-Posttest Design

In a school situation, it is sometimes practically not possible to upset class schedules, to gather subjects for obtaining a sufficiently large sample or to reorganize classes in order to employ randomization procedures for getting equivalent control and experimental groups. Under these circumstances, therefore, an experimenter may use pre-assembled groups, such as intact classes, for framing experimental and control groups. The pre-assembled groups are selected and are administered pretest. The pretest scores are analysed to show that the means and standard deviations of the two groups do not differ significantly. If the pretest scores for the groups are not equivalent, the experimenter may proceed with the conduct of the experiment by using the technique of analysis of co-variance to compensate for this lack of equivalency between the groups. Once the two groups are obtained, it is advisable to use a random procedure to determine which group is to be assigned to experimental treatment and which one to the controlled condition. After determining the groups the experimental treatment is administered to the experimenta group and then the posttest is given to both the groups. The difference between the pre-and posttest scores are compared with the help of appropriate statistical test to ascertain the effect of the independent variable (X).

PARADIGM FOR DESIGN 9: Non-randomized Control-group, Pretest-Posttest Design

Group	*Pretest*	*Independent variable*	*Posttest*
Experimental	T_1	Experimental treatment	T_2
Control	T_1	Controlled condition	T_2

Advantages

1. The reactive effects of experimentation are more easily controlled than in Design 5. When the preassembled groups are used, subjects are less aware of the fact that they are subjected to the experimental treatment than when the subjects are drawn from class through randomization and put into experimental sessions.

2. The experiments using this design are conveniently conducted in the school situations.

Limitations

1. The selection of subjects for the experimental and control groups may result in interaction effect between selection and certain extraneous variables like selection and maturation, and selection and history, etc. that could be mistakenly attributed to the effect of independent variable.

2. Statistical regression is a major threat to the internal validity of this design.

3. The sources of threats to external validity in this design are the same as were discussed in the Design 5.

Design 10: Counterbalanced Design

When the random assignment of subjects to experimental and control groups is not possible, the counterbalanced design may be used. This design also uses intact classes and overcome some of the weaknesses in Design 9 by rotating the groups at periodic intervals during the experimentation. This design is also known as rotation-group design, crossover design or switchover design.

In a counterbalanced design each group of subjects is exposed to each experimental treatment (X) at different times during the experiment. During the first exposure to the experimental treatment, for example, if group 1 is exposed to treatment X_1 and group 2 to X_2, the second time group 1 is exposed to X_2 and group 2 to X_1. The sums of the scores (X_1+X_1) and (X_2+X_2) of the two groups are compared with the help of an appropriate statistical test to ascertain the effect of the experimental treatment. The significant difference between the sums (X_1+X_1) and (X_2+X_2) cannot be interpreted as the product of the initial difference between the two groups, for each group affects each X exactly once; nor can the difference be interpreted as the product of the difference between order of testing, for each time of testing affects each X exactly once.

Illustration

Suppose a classroom teacher wants to compare the effectiveness of Method *A* and Method *B* on achievement in social studies of sixth grade students. The teacher could choose two of his classes (group 1 and group 2) and two units (unit 1 and unit 2) of social studies equivalent in the complexity and difficulty of the concepts involved. During first replication of experiment, group 1 is taught unit 1 by Method *A* and group 2 is taught the same unit by Method *B*. At the end of instruction, an achievement test on unit 1 is administered to both groups. Then during the second replication, group 1 is taught unit 2 by Method *B*, and group 2 is taught the same unit by Method *A*. An achievement test on unit 2 is administered to both the groups at the end of the instruction.

PARADIGM FOR DESIGN 10: The Counterbalanced Design

	Experimental Treatments	
Replication	*Method A*	*Method B*
Unit 1	Group 1	Group 2
Unit 2	Group 2	Group 1
	Column mean	Column mean

At the end of the experiment, the column means are computed to get the mean achievement for the two groups when taught by the method shown by the column heading. The column mean scores are compared to ascertain the effectiveness of the methods upon achievement in social studies.

Advantages

1. This design overcomes the weaknesses of Design 9. When intact groups are used, rotation of groups provides an opportunity to eliminate any differences that might exist between the groups.
2. Since all the groups are exposed to all the treatments, the results obtained cannot be attributed to preexisting differences in the subjects.

Limitations

1. There is carry-over effect of the groups from one treatment to the next. Therefore, this design should be used only when the experimental treatments are such that administration of one treatment on a group will have no effect on the next treatment.
2. It is not always possible to have equivalent learning material which the experimenter wants to use in the various replications.

interval validity of the design.

18. Define external validity of the design.
19. List and describe the types of external validity of the design.
20. Define population validity of the design.
21. Define ecological validity of the design.
22. List and describe the factors which contribute towards achieving ecological validity of the design.
23. Define the terms: experimenter, subject, independent variable, dependent variable, control group, experimental group, pretest and posttest.
24. List and describe various types of experimental designs.
25. Describe the nature of preexperimental designs.
26. List the various types of preexperimental Designs.
27. Describe the nature of one Group Pretest-Posttest Design.
28. Illustrate the use of one Group Pretest-Posttest Design.
29. List and describe the limitations of one Group Pretest-Posttest Design.
30. Describe the nature of two Groups, Static-Design.
31. Illustrate the use of two Groups, Static-Design.
32. List and describe the limitations of two Groups, Static-Design.
33. Describe the nature of true Experimental Designs.
34. List the various types of true Experimental Designs.
35. Describe the nature of two Groups, Randomized Subjects, Posttest—only Design.
36. Illustrate the use of two Groups, Randomized Subjects, Posttest—only Design.
37. List and describe the advantages of two Groups, Randomized Subjects, Posttest—only Design.
38. List and describe the limitations of two Groups, Randomized Subjects, Posttest—only Design.
39. Describe the nature of two Groups, Randomized Matched subjects, Posttest—only Design.
40. Illustrate the use of two Groups, Randomized Matched Subjects, Posttest—only Design.
41. List and describe the advantages of two Groups, Randomized Matched Subjects, Posttest—only Design.
42. List and describe the limitations of two Groups, Randomized Matched Subjects, Posttest—only Design.
43. Describe the nature of Randomized Groups, Pretest-Posttest Design.
44. Illustrate the use of Randomized Groups, Pretest-Posttest Design.
45. List and describe the advantages of Randomized Groups, Pretest-Posttest Design.
46. List and describe the limitations of Randomized Groups, Pretest-Posttest Design.
47. Describe the nature of Randomized Solomon Three-Group Design.
48. Illustrate the use of Randomized Solomon Three-Group Design.
49. List and describe the advantages of Randomized Solomon Three-Group Design.
50. List and describe the limitations of Randomized Solomon Three-Group Design.
51. Describe the nature of Randomized Solomon Four-Group Design.
52. Illustrate the use of Randomized Solomon Four-Group Design.

53. List and describe the advantages of Randomized Solomon Four-Group Design.
54. List and describe the limitations of Randomized Solomon Four-Group Design.
55. Describe the nature of Factorial Designs.
56. Describe the nature of Simple Factorial Design of 2 by 2 (2×2).
57. Illustrate the use of Simple Factorial Design of 2 by 2 (2×2).
58. List and describe the advantages of Factorial Designs.
59. List and describe the limitations of Factorial Designs.
60. Describe the nature of Quasi experimental Designs.
61. List the various types of Quasi Experimental Designs.
62. Describe the nature of Non-randomized control Group, Pretest-Posttest Design.
63. Illustrate the use of Non-randomized control Group, Pretest-Posttest Design.
64. List and describe the advantages of Non-randomized control Group, Pretest-Posttest Design.
65. List and describe the limitations of Non-randomized Control Group, Pretest-Posttest Design.
66. Describe the nature of counterbalanced Design.
67. Illustrate the use of counterbalanced Design.
68. List and describe the advantages of counterbalanced Design.
69. List and describe the limitations of Counterbalanced Design.
70. Describe the nature of Time-Series Designs.
71. List the various types of Time-Series Designs.
72. Describe the nature of One-Group Time-Series Design.
73. Illustrate the use of One-Group Time-Series Design.
74. List and describe the advantages of One-Group Time-Series Design.
75. List and describe the limitations of One-Group Time-Series Design.
76. Describe the nature of Control-Group Time-Series Design.
77. Illustrate the use of Control-Group Time-Series Design.
78. List and describe the advantage of Control-Group Time-Series Design.
79. List and describe the limitations of Control-Group Time-Series Design.

REFERENCES

Aaron, P.G. (1968). " Research Needs in Educational Psychology", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia.Kurukshetra University Books and Stationery Shop, pp. 164-170.

Adaval, S. B. ed.(1968). *Third Indian Year Book of Education: Educational Research*. New Delhi : National Council of Educational Research and Training.

Aitken, A.C. (1937). " The Evaluation of a Certain Triple Product Matrix" *in Statistical Analysis in Psychology and Education*. George A. Ferguson, Tokyo: McGraw Hill Kogakusha, 1976, pp. 453-468.

Anderson, J. Berry, H. Durston, and Millicent Poole, (1971). *Thesis and Assignment Writing*. New Delhi: Wiley Eastern Limited.

Ansari, N. (1965). *Economics of Education in Indian Context – An Agenda for Research*. New Delhi: National Council for Educational Research and Training.

Ary, D., Lucy C. Jacobs., and Asghar Razavieh. (1972). *Introduction to Research in Education*. New York: Holt, Rinehart and Winston, Inc.

Babbie, E.R. (1973). *Survey Research Methods*. Belmont, Calif: Wadsworth Publishing Co.

Banker, Robbert L. and Richard E. Schutz. (1972). *Instructional Product Research*. New York: American Book Co.

Barr, A.S., Robert A. Davis, and Plamer O. Johnson. (1953). *Educational Research and Appraisal*. Chicago: J.B. Lippincott Co.

Barzon, J. and Henry F. Graff. (1970). *The Modern Researcher*. New York: Harcourt Brace Jovanovich.

Becker, G.S. (1965). *Human Capital: A Theoretical and Empirical Anaysis, With Special Reference to Education*. Princeton, N. J.: Princeton University Press.

Beggs, Donald L. and Ernest L. Lewis. (1975). *Measurement and Evaluation in the Schools*. Boston: Houghton Mifflin Company.

Best, John W. (1977). *Research in Education*. New Delhi: Prentice-Hall of India Private Limited.

Bhat, R.S. (1968). "Research Needs in the Philosophy of Education", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 128-138.

Bird, C. (1940). *Social Psvchology*. New York: Appleton—Century—-Crofts.

Bloom, Behjamin S., *et al* eds. (1956). *Taxonomy of Educational Objectives 1: Cognitive Domain*. New York: David Makay Co., Inc.

Borg, Walter and Meredith D. Gill. (1971). *Educational Research: An Intorduction*. New York: David Mackay Co. Inc.

Borrowman, Marle L. (1960). " History of Education" in

Encyclopaedia of Educational Research, ed. Chester W. Harris, New York: The Macmillan Company.

Brandt, Richard M. (1972). *Studying Behaviour in Natural Settings.* New York: Holt, Rinehart and Winston.

Bracht, G. and G. Glass. (1968). " The External Validity of Experiments", *American Educational Research Journal,* 5, pp. 437-474.

Brickman, William W. (1949). *A Guide to Research in Educational History.* New York: New York University Bookstore.

Buch, M.B. ed. (1973). *A Survey of Research in Education.* Baroda: Centre of Advanced Study in Education, M.S.University.

Buch, M.B. ed. (1979). *Second Survey of Research in Education.* Baroda: Society for Educational Research and Development.

Buch, M.B. ed. (1587). *Third Survey of Research in Education.* New Delhi: NCERT.

Buros, Oscar, K. ed. (1936 to Present). *The Mental Measurements Yearbooks.* Highland Park, New Jersey: Gryphon Press.

Buswell, Guy T. (1966). *Training for Educational Research.* Berkeley, Calif. Centre for the Study of Higher Education, University of California.

Campbell, Donald T. and Jullian C. Stanley. (1963). "Experimental and Quasi-Experimental Designs for Research on Teaching", in *Handbook of Research on Teaching,* N.L. Gage, ed. Chicago: Rand McNally and Company, 1963, pp. 175-176.

Campbell, W.G. (1969). *Form and Style in Thesis Writing.* Boston: Houghton Mifflin Company.

Carr, Edward H. (1962). *What is History ?* New York: Alfred A. Knopt.

Chatterji, S. and M. Mukerjee. (1968). "Research Needs in Educational Measurement and Evaluation", in *Research Needs in the Study of Education.* eds. Uday Shanker and S.P Ahluwalia. Kurukshetra; Kurukshetra University Books and Stationery Shop, pp. 198-209.

Cochran, W.G. and G. M. Cox. (1950). *Experimental Designs.* New York: John Wiley and Sons, Inc.

Cooper, Dan, H. (1946). "Contributions of School Surveys of Educational Administration" in *Encyclopaedia of Educational Research.* C.W. Harris, ed. New York : The Macmillan Company. 1960, p. 1212.

Corey, Stephen M. (1953). *Action Research to Improve School Practices.* New York: Teachers College, Columbia University.

Corey, Stephen M. and J.K. Shukla. (1962). *Parctical Classroom Research By Teachers.* New Delhi: The Manager of Publications.

Cronbach, Lee J. (1964). *Essentials of Psychological Testing.* New York: Harper and Row, International Education.

Culbertson, Jack A. and Stephen p. Hencley, eds. (1963). *Educational Research: New Perspectives.* Danville, Ill.: Interstate Publishers and

Printers.

Dececco, John P. and William R. Crawford. (1977). *The Psychology of Learning and Instruction*. New Delhi: Prentice-Hall of India Private Ltd., p. 67.

Deo, Pratibha, (1968) "Research Needs in Educational Psychology", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia. Kurukshetra.: Kurukshetra University Books and Stationery Shop, pp.171-174.

Dewey, John, (1933). *How We Think ?*. Boston: D.C. Heath and Company.

Dugdale, K. (1967). *A Manual on Writing Research*. Bloomington, Indiana: Indiana University Bookstore.

Dugdale, K. (1967). *A Manual of Form for Theses and Term Papers*. Bloomington, Ind: Indiana University Bookstore.

Ebel, R.L. ed. (1969). *Encyclopaedia of Educational Research*. London: The Macmillan Company.

Edwards, A.L. and K.C. Kenny (1946). "A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction", *J. Appl. Psychol.*, 30, pp. 72-83.

Edwards, A.L. and F.P. Kilpatrick. (1948). "A Technique for the Construction of Attitude Scales". *J. Appl. Psychol.* 32, pp. 374-384.

Edwards, A.L. (1957). *Techniques of Attitude Scale Construction*, New York: Appleton-Century Crofts, Inc.

Edwards, Allen L. (1971). *Experimental Design in Psychological Research*. New Delhi: Amerind Publishing Co. Private Ltd. Indian Edition.

Englehart, Max D. (1972). *Methods of Educational Research*. Chicago: Rand McNally and Company.

Ferguson, L.W. (1939). "The Requirements of an Adequate Attitude Scale", *Psychol. Bull.*, 36; pp. 665-673.

Festinger, L., and D. Katz eds. (1976). *Research Methods in Behavioural Sciences*. New Delhi: Amerind Publishing Co. Pvt. Ltd. Fourth Indian Reprint.

Fisher, R. A. (1935). *The Design of Experiments*. London: Oliver and Boyd, pp. 200-201.

Flanders, N. (1969). "Teacher Effectiveness", in *Encyclopaedia of Educational Research*. ed. R. L. Ebel. London: The Macmillan Co., pp 1423-37.

Fox, David J. (1969). *The Research Process in Education*. New York: Holt, Rinehart and Winston.

Freeman, Frank S. (1965). *Theory and Practice of Psychological Testing.* New Delhi: Oxford and IBH Publishing Co. Indian Edition.

Gage, N.L. (1963). *Handbook of Research on Teaching.* Chicago: Rand McNally and Company.

Galfo, Armand J. (1975). *Interpreting Educational Research*. Dubuque, Iowa: William C. Brown Co.

Garraghan, Gilbert J. (1946). *A Guide to Historical Method.* New York: Fordham University Press.

Garrett, H.E. (1962). *Statistics in Psychology and Education.* Bombay: Allied Pacific Private Limited.

Gephart, William J. and Robert B. Ingle. (1969). *Educational Research: Selected Readings.* Columbus, Obio: Charles E. Merril Publishing Co.

Gesell, A., *et al* (1939). *Biographies of Child Development.* New York: Hoeber.

Glaser, Robert, (1962). "Psychology and Instructional Technology" in *Training, Research and Education.* ed. Robert Glaser, Pittsburgh: University of Pittsburgh Press, pp.1-30.

Glaser, Robert *et al* (1966). *Organization for Research and Development in Education,* Bloomington, Ind.:Phi Delta Kappa.

Good, Carter V., A.S. Barr and Douglas E. Scates (1941). *Methodology of Educational Research.* New York: Appleton-Century Crofts, Inc.

Good, Carter V., ed, (1959). *Dictionary of Education.* New York: McGraw-Hill Company, p.149

Good, Carter V. (1966). *Essentials of Educational Research: Methodology and Design.* New York: Appleton-Century Crofts.

Goode, William J. and Paul K. Halt. (1952). *Methods in Social Research.* New York: McGraw-Hill Book Company.

Gordon, Ira J. (1965). *Human Development: Readings in Research.* Glenview, Ill.: Scott Foresman and Company.

Gottschalk, Louis R. (1951). *Understanding History.* New York: Alfred A. Knopf.

Gottschalk, Louis, R. ed. (1963). *Generalization in Writing of History.* Chicago: University of Chicago Press, 1963.

Government of India. Ministry of Education. *Report of the Secondary Education Commission.* (1952-1953), New Delhi

Government of India, Ministry of Education. *Education and National Development. Report of the Education Commission* (1964-66), New Delhi.

Green, P.E. (1972). *Applied and Multidimensional Scaling: A Comparison of Approaches and Algorithms.* New York: McGraw-Hill.

Guilford, J.P. (1954). *Psychometric Methods.* New Delhi: Tata McGraw-Hill Publishing Co. Ltd.

Guilford, J.P. (1965). *Fundamental Statistice in Psychology and Education.* New York McGraw-Hill Book Company.

Guilford, J.P. (1967). *The Nature of Human Intelligence* New York: McGraw-Hill.

Hall, O.M. (1934). "Attitude and Unemployment", *Arch. Psychol.,* No. 165.

Hardwick, Curtis D. and Lewis F. Petrinovich. (1975), *Understanding Research in the Social Sciences*. Philadelphia : W.B. Saunders Co.

Hartshorne, H. and M.A. May. (1929). *Studies in Service and Self Control*. New York : Macmillan Company.

Harris Chester W., ed. (1960). *Encyclopaedia of Educational Research*. New York : The Macmillan Company.

Harris, S.E. (1962). *Higher Education : Resources and Finance*. New York: McGraw-Hill

Harper, Edwin A. (1968). "Research Needs in Educational and Vocational Guidance", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra : Kurukshetra University Books and Stationery Shop, pp. 182-197.

Hayman, John L. (1968). *Research in Education*. Columbus, Ohio : Charles E. Merrill Publishing Company.

Hillway, Tyrus. (1964). *Introduction to Research*. Boston : Houghton Mifflin Company.

Hokett, Homer C. (1955). *The Critical Method in Historical Research and Writing*. New York : The Macmillan Company.

Honzik, M.P., J.W. Macfarland and L. Allen (1949). "The Stability of Mental Test Performance between Two and Eighteen Years," *J. Expt. Educ.*, 17, pp. 309-324.

Johnson, Palmer O. (1961). *Statistical Method in Research*. Bombay: Asia Publishing House.

Kahn, Robert L. and Charles F. Cannel. (1957). *The Dynamics of Interviewing*. New York : John Wiley and Sons.

Kaplan, Abraham. (1964). *The Conduct of Inquiry : Methodology for Behavioural Science*. San Francisco: Chandler Publishing Co.

Kelley, T.L. *et al* (1953). *Stanford Achievement Test*. New York: Harcourt, Brace and World Inc.

Kempfer, Homer, (1960). "Tools for Evaluation", *Indian Jorunal of Adult Education*. Vol. XXI, No. 7.

Keppel, G, (1973). *Design and Analysis : A Researcher's Handbook*. Englewood Cliffs, N. J.: Prentice-Hall.

Kerlinger, Fred N. ed. (1973). *Review of Research in Education* 1. Itasca, Ill.: F.E. Peacock, Publishers.

Kerlinger, Fred N. and John B. Carroll eds. (1974). *Review of Research in Education II*. Itasca, Ill.: F.E. Peacock, Publishers.

Kerlinger, Fred N. et al (1975). *Review of Research in Education III. Itasca, III.:* F.E. Peacock, Publishers.

Kerlinger, Fred N. (1978). *Foundations of Behavioural Research*. New York : Holt, Rinehart and Winston, Inc. First Indian Reprint.

Kershaw, J.A. and R.N. McKean, (1959). *Systems Analysis and Education*, New York : The Rand Corporation.

King, Edward J. (1965). *Comparative Education*. Vol. 1, No. 3, p. 157.

King, Edmund J.(1965). *World Perspectives in Education*. London:

Methuen and Co. p. 358.

Koul, Lokesh (1974). "Examination Reform at the School Level", *NIE Journal* Vol VIII, No. 4, p. 6-9.

Koul, Lokesh (1974a) "Examination Reforms", *Kurukshetra University Research Journal.* Vol. VIII, Part 1 and 2, pp. 1-8.

Koul, Lokesh. (1976). "Implications of Work Experience", *H.P. Board Journal of Education.* Vol. I, pp. 26-28.

Koul, Lokesh (1977). "Significant Innovations in Indian Examination System : A Review", *Himshiksha* (Special Issue on Examination Reform). Vol. VII, NO. 2, pp.1-4.

Koul, Lokesh. (1978). "Non-Formal Education: Why and How ?" *The Education Quarterly* Vol. XXIX, No. 4, pp. 25-27.

Koul, Lokesh, (1979). "Implication of Adult Education", *The Education Quarterly* Vol. XXX, No. 4, pp. 9-11.

Koul, Lokesh. (1979 a). "Priorities in Adult Education", *Workers Education* pp. 30-33.

Koul, Lokesh (1985). "Validity of the Experimental Designs in Educational Research", in *Quality Control in Educational Research* eds. S.K. Pal and P.C. Saxena. New Delhi; Metropolitan, 1985.pp. 120-140.

Koul, Lokesh (1985). "Choosing an Appropriate Statistical Technique for Analysis of Data in Experimental Research", in *Quality Control in Educational Research,* eds., S.K. Pal and P.C. Saxena, New Delhi : Metropolitan, 1985, pp. 150-163.

Kuder, G.F. (1954). "Expected Developments in Interest and Personality Inventories", *Educational and Psychological Measurement.* Vol. XXIV, No. 2.

Kulkarni, S.S., Mohan Lal, and C.A.S. Naidu. (1968). *All India Survey in Mathematics Achievement: Primary, Middle and High School Levels.* New Delhi : NCERT.

Kundu, C.L. (1977). *Personality Development : A Critique of Indian Studies.* Kurukshetra : Vishal Publications.

Kunda, C.L. (1980). *Rorschach Psychodiagnostics: An Indian Case Study.* Kurukshetra: Vishal Publications.

Kuppuswamy, B. (1968). "Research Needs in Educational Psychology", in *Research Needs in the Study of Education.* eds. Uday Shanker and S.P. Ahluwalia. Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 155-163.

Lehmann, Irvin J. and William A. Mehrens, eds. (1971). *Educational Research: Readings in Focus.* New York: Holt, Rinehard and Winston, Inc.

Likert, R. (1932). "A Technique for the Measurement of Attitudes", *Arch. Psychol.* No. 140.

Lovell, K., and K.S. Lawson (1973). *Understanding Research in Education.* London : University of London Press.

Lulla, B.P. (1968). "Research Needs in School Curriculum", in *Research Needs in the Study of Education* eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop. pp. 139-149.

Mill, John S. (1846). *A System of Logic*. New York: Harper and Row, P. 222.

Mittal, R.L. (1979). *Advances in the Liberary and Information Science*. Delhi : Moon Light Publishers and Distriburtors.

Mitra, Shib K. (1958). "Problems in Transition from School to College", *Indian University Administration*. New Delhi: Ministry of Education, Government of India.

Mitra, Shib K.(1968). "Research Needs in Test Development", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 210-216.

Monroe, W.S. and M.D. Engelhart. (1928). "The Techniques of Educational Research", Urbana, Illinois: Bureau of Educational Rsearch, University of Illions, Bulletin No. 38, p. 14.

Monroe, Watlers S., ed. (1950). *Encyclopaedia of Educational Research*. New York: The Macmillan Company.

Mouly, Geroge J. (1963). *The Science of Educational Research*. New Delhi: Eurasia Publishing House Private Limited.

Murray, H.A.(1938). *Explorations in Personality*. New York: Oxford Unviersity Press.

Naidu, P.S. (1968). "Research Needs in the Philosophy of Education", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia. Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 122-127.

Naik, J.P. (1968). "Research Needs in Elementary Education", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia (paper reproduced from the *Indian Year Book of Education: Second Year Book: Elementary Education)*, pp. 234-235.

NCERT. (1975). *Research Schemes*. New Delhi: National Council of Educational Research and Training.

Neale, J.M. and R.M. Liebert (1973). *Science and Behaviour: An Introduction to Mehtods of Research*, Englewood Cliffs, N.J.: Prentice-Hall.

Nevins, A. (1963). *The Gateway to History*. Chicago: Qudrangle Books.

Nystrom, G.H. (1933). "The Measurement of Fillipino Attitudes towards America by the use of Thurstone Technique." *J.Soc. Psychol.* 4, pp.249-252.

Oad, L.K. (1968). "Research Needs in Comparative Education", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop. pp. 56-65.

Pandey, K.P. (1975). "Non-Formal Education and Community Colleges" Paper Presented in Seminar on "Job Oriented Education and Community Colleges" Under the sponsorship of *USIS*, New Delhi.

Pandit, H.N.(1967). *Monograph on Plan of Studies in Economics of Education*. New Delhi: National Council of Educational Research and Training.

Pareek, Udai.(1967). *Directory of Behavioural Science Research*. Delhi: Macmillan Company.

Raina, M.K. (1986). *Education of the Left and the Right : Implications of Hemispheric Specialisation*. New Delhi, Allied Publishers.

Raina, M.K. BEd. (1986) *Biochemicals and Behiviour*, New Delhi: National Council of Educational Research and Training.

Ray, William S. (1960). *An Introduction to Experimental Design*. New York: The Macmillan Company.

Reller. Theodore L.(1942). "Shall we have a Status, Deliberation, or implementative study of our schools?" in *Encyclopaedai of Educational Research*. C.W. Harris, ed. New York. The Macmillan Company, 1960, p. 1212.

Rosander, A.C. (1936). "The Spearman-- Brown Formaula in Atttitude Scale Construction", *J. Expt. Psychol*, 19, pp. 486-495

Rundquist, E.A., and R.F. Sletto, (1936). *Personality in the Depression.* Minneapolis: University of Minnesota Press.

Runkla, Phillip and Joselh McGrath. (1972). *Research in Human Behaviour*. New York: Holt, Rinehart and Winston.

Rugg. H.O. (1921). "Is the Rating of Human Character Practicable?" *J. Educ. Psychol. in Psychometric Methods* J.P. Guilford New Delhi: Tata McGraw Hill Publishing Co. Ltd. 1954, p. 279.

Rusk, Robert R.(1960). *An Outline of Experimental Education*, New York: The Macmillan Company.

Sax, Gilbert.(1968). *Empirical Foundation of Educational Research.* Englewood Cliffs, N.J.: Prentice-Hall.

Scannell Dale P. and D.P. Trecy (1975). *Testing and Measurement in the Classroom* Boston: Houghton Mifflin Company.

Schultz, T.W. (1963). *The Economic Value of Education*. New York: Columbia University Press.

Selltiz, claire, Marie Jahoda, Morton Deutch, and Stuart W. Cook. (1961). *Research Methods in Social Relations*, New York: Holt, Rinehart and Winston.

Shanker, U. (1968). "Research Needs in the Education of Exceptional Children", in *Research Needs in the Study of Education*, eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukeshetra University Books and Stationery Shop, p. 175.

Sharma, R.C. (1968). "Research Needs in Educational Administration," in *Research Needs in the Study of Education*. eds. Uday

Shanker and S.P. Ahulwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 94-104.

Siegel, S.(1956). *Nonparametric Statistics for the Behavioural Sciences Tokyo:* McGraw-Hill Kogakusha, Ltd.

Singh, H.(1968). "Research Needs in Economics of Education", in *Research Needs in Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 217-225.

Singh, Kulwinder (1982). "Problems in Measurement of Creative Thinking". *ISPT Journal of Education and Psychology*. Vol. VI, No. 2. pp.22-25.

Singh, R.P. (1968). "Research Needs in Comparative Education", in *Research Needs in the Study of Education*. eds. Uday Shanker and S.P. Ahluwalia, Kurukshetra: Kurukshetra University Books and Stationery Shop, pp. 66-71.

Silvey, J (1975). *Deciphering Data: The Analysis of Social Surveys.* London, England: Longman.

Sisson, E.E. (1948). "Forced Choice: The New Army Rating", *Person. Psychol.*, 1, pp. 365-381.

Stevens, S.S.(1951). *Handbook of Experimental Psychology*. New York: John Wiley.

Sukhla, S.P., P.V. Mehrotra, and R.N. Mehrotra (1966). *Elements of Educational Research*. New Delhi: Allied Publishers Private Limited.

Symonds, P.M.(1925). "Notes on Rating", *J. Appl. Psychol. in Psychometric Mehtods* J.P. Guilford. New Delhi: Tata McGraw-Hill Publishing Co. Ltd. 1954, p. 279.

Taylor, E.K., and R.W. Wherry. (1951). "A Study of Leniency in Two Rating Systems", *Person Psychol.* 4, pp. 39-47.

Terman, Lewis, M., and M.H. Oden 1947. *The Gifted Child Grows up: Twenty five Years Follow up Studies of a Superior Group* Vol. IV (Stanford clif: Stanford University Press).

Thomson, G. (1960). *The Factorial Analysis of Human Ability*. London: University of London Press, Ltd.

Thorndike, Robert L. and E. Hagen (1970). *Measurement and Evaluation in Psychology and Education* New Delhi: Wiley Eastern Private Ltd.

Thorp, William L. (1962). "101 Questions for Investigatons" in *Economics of Higher Education*, ed. Selma J. Mushkin, Washington, D.C.: U.S. Department of Health, Education and Welfare.

Thurstone, L.L., and E.J. Chave. (1929). *The Measurement of Attitude.* Chicago: The University of Chicago Press.

Thurstone, L.L. (1946). "Comments". *Amer. J. Social*, 52, pp. 39-50.

Thurstone, L.L. (1947). *Multiple-Factor Analysis*. Chicago: The University of Chicago Press.

Torrance, E.P.(1966). *Torrance Tests of Creative Thinking.* Princeton, New Jersey: Personnel Press.

Travers, Robert M.W. (1958). *An Intorduction to Educational Research.* New York: The Macmillan Company.

Travers, Robert M.W. (1969). *An Introduction To Educational Research.* New York: The Macmillan Company .

Travers, Robert M.W, (1973). *The Handbook of Research on Teaching.* Chicago: Rand McNally and Co.

Travers, Robert M.W. (1978). *An Introduction to Educational Research.* New York: Macmillan Publishing Company, Inc.

Tuckman, Bruce W. (1972). *Conducting Educational Research.* New York: Harcourt Brace Jovanovich.

Tuckman, B.W. (1975). *Measuring Educational Outcomes: Fundamental of Teaching.* New York: Harcourt Brace Jovanovic Inc.

Tuckman, B.W. (1979). *Evaluating Instructional Programmes.* London: Alleyn and Bacon, Inc.

Turney, B and George, Robb (1971). *Research in Education: An Introduction. Illinois:* The Dryden Press Inc.

Turabian, Kate L. (1967). *A Manual For Writers of Term Papers, Theses, and Dissertations.* Chicago: The University of Chicago Press.

Tyrus, Hillway. (1964). *Introduction to Research. Boston:* Houghton Mifflin Company.

Uhbrock, R.S. (1934). "Attitude of 4430 Employees", *J. Soc. Psychol.* 5, pp. 365-377

UNESCO. (1951). "Curriculum Research and Revision" *Bulletin No. 28* Paris: UNESCO., pp. 27-28.

UNESCO (1972). *Learning To Be: The World of Education Today and Tomorrow.* New Delhi: Sterling Publishers Private Limited.

Van Dalen, Deobold B.(1973). *Understanding Educational Research.* New York: McGraw-Hill Book Company.

Varma, M.(1965). *An Introduction to Educational and Psychological Research.* New Delhi: Asia Publishing House.

Vernon, P.E.(1962). *Personality Tests and Assessment.* London: Methuen and Co.

Wandit, Edwin. (1965). *A Cross Section of Educational Research.* New York: David Mckay Co.

Wang, K.A. (1932). "Suggested Criteria for Writing Attitude Statements", *J. Soc. Psychol.* 3, pp. 367-373.

Whitney, Frederick L. (1964). *The Elements of Research.* Englewoods Cliffs, New Jersey: Prentice-Hall, Inc.

Wiersma, William,(1975). *Research Methods in Education.* Itasca, III. F.E. Peacock Publishers.

APPENDIX

TABLE A: Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distances on the base line between the mean and successive points laid off from the mean in units of standard deviation.

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3290	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4920	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3.0	4986.5	4986.9	4987.4	4987.8	4988.2	4988.6	4988.9	4989.3	4989.7	4990.0
3.1	4990.3	4990.6	4991.0	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	4992.9
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

TABLE B: Ordinates of the normal probability curve expressed as fractional parts of the mean ordinate, Y.

$\frac{x}{\sigma}$	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0.0	100000	99995	99980	99955	99920	99875	99820	99755	99685	99596
0.1	99501	99396	99283	99158	99025	98881	98728	98565	98393	98211
0.2	98020	97819	97609	97390	97161	96923	96676	96420	96156	95882
0.3	95600	95309	95010	94702	94387	94055	93723	93382	93024	92677
0.4	92312	91399	91558	91169	90774	90371	89961	89543	89119	88688
0.5	88250	87805	87353	86896	86432	85962	85488	85006	84519	84060
0.6	83527	83023	82514	82010	81481	80957	80429	79896	79359	78817
0.7	78270	77721	77167	76610	76048	75484	74916	74342	73769	73193
0.8	72615	72033	71448	70861	70272	69681	69087	68493	67896	67298
0.9	66689	66097	65494	64891	64287	63683	63077	62472	61865	61259
1.0	60653	60047	59440	58834	58228	57623	57017	56414	55810	55209
1.1	54607	54007	53409	52812	52214	51620	51027	50437	49848	49260
1.2	48675	48092	47511	46933	46357	45783	45212	44644	44078	43516
1.3	42956	42399	41845	41294	40747	40202	39661	39123	38569	38058
1.4	37531	37007	36487	35971	35459	34950	34445	33944	33447	32954
1.5	32465	31980	31500	31023	30550	30082	29618	29158	28702	28251
1.6	27804	27361	26923	26489	26059	25634	25213	24797	24385	23978
1.7	23575	23176	22782	22392	22008	21627	21251	20879	20511	20148
1.8	19790	19436	19086	18741	18400	18064	17732	17404	17081	16762
1.9	16448	16137	15831	15530	15232	14939	14650	14364	14083	13806
2.0	13534	13265	13000	12740	12483	12230	11981	11737	11496	11259
2.1	11025	10795	10570	10347	10129	09914	09702	09495	09290	09090
2.2	08892	08698	08507	08320	08136	07956	07778	07604	07433	07265
2.3	07100	06939	06780	06624	06471	06321	06174	06029	05888	05750
2.4	05614	05481	05350	05222	05096	04973	04852	04734	04618	04505
2.5	04394	04285	04179	04074	03972	03873	03775	03680	03586	03494
2.6	03405	03317	03232	03148	03066	02986	02908	02831	02757	02684
2.7	02612	02542	02474	02408	02343	02280	02218	02157	02098	02040
2.8	01984	01929	01876	01823	01772	01723	01674	01627	01581	01536
2.9	01492	01449	01408	01367	01328	01288	01252	01215	01179	01145
3.0	01111	00819	00598	00432	00309	00219	00153	00106	00073	00050
4.0	00034	00022	00015	00010	00006	00004	00003	00002	00001	00001
5.0	00000									

TABLE C: Table of Critical Values of t.

Degrees of Freedom	*Level of Significance* *0.10*	 *0.05*	 *0.02*	 *0.01*
1	6.34	12.71	31.82	63.66
2	2.92	4.30	6.96	9.62
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.90	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.06
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.72	2.09	2.53	2.84
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.50	2.81
24	1.71	2.06	2.49	2.80
25	1.71	2.06	2.48	2.79
26	1.71	2.06	2.48	2.78
27	1.70	2.05	2.47	2.77
28	1.70	2.05	2.47	2.76
29	1.70	2.04	2.46	2.76
30	1.70	2.04	2.46	2.75
35	1.69	2.03	2.44	2.72
40	1.68	2.02	2.42	2.71
45	1.68	2.02	2.41	2.69
50	1.68	2.01	2.40	2.68
60	1.67	2.00	2.39	2.66
70	1.67	2.00	2.38	2.65
80	1.66	1.99	2.38	2.64
90	1.66	1.99	2.37	2.63
100	1.66	1.98	2.36	2.63
125	1.66	1.98	2.36	2.62
150	1.66	1.98	2.35	2.61
200	1.65	1.97	2.35	2.60
300	1.65	1.97	2.34	2.59
400	1.65	1.97	2.34	2.59
500	1.65	1.96	2.33	2.59
1000	1.65	1.96	2.33	2.58
∞	1.65	1.96	2.33	2.58

TABLE D: F-ratios for .05 (roman) and .01 (bold face) Levels of Significance

Degrees of freedom for smaller mean square	*Degrees of freedom for greater mean square*									
	1	*2*	*3*	*4*	*5*	*6*	*8*	*12*	*24*	∞
1	161.45 **4052.10**	199.50 **4999.03**	215.72 **5403.49**	224.57 **5625.14**	230.17 **5764.08**	233.97 **5859.39**	238.89 **5981.34**	243.91 **6105.83**	249.04 **6234.16**	254.32 **6366.48**
2	18.51 **98.49**	19.00 **99.01**	19.16 **99.17**	19.25 **99.25**	19.30 **99.30**	19.33 **99.33**	19.37 **99.36**	19.41 **99.42**	19.45 **99.46**	19.50 **99.50**
3	10.13 **34.12**	9.55 **30.81**	9.28 **29.46**	9.12 **28.71**	9.01 **28.24**	8.94 **27.91**	8.84 **27.49**	8.74 **27.05**	8.64 **26.60**	8.53 **26.12**
4	7.71 **21.20**	6.94 **18.00**	6.59 **16.69**	6.39 **15.98**	6.26 **15.52**	6.16 **15.21**	6.04 **14.80**	5.91 **14.37**	5.77 **13.93**	5.63 **13.46**
5	6.61 **16.26**	5.79 **13.27**	5.41 **12.06**	5.19 **11.39**	5.05 **10.97**	4.95 **10.67**	4.82 **10.27**	4.68 **9.89**	4.53 **9.47**	4.36 **9.02**
6	5.99 **13.74**	5.14 **10.92**	4.76 **9.78**	4.53 **9.15**	4.39 **8.75**	4.28 **8.47**	4.15 **8.10**	4.00 **7.72**	3.84 **7.31**	3.67 **6.88**
7	5.59 **12.25**	4.74 **9.55**	4.35 **8.45**	4.12 **7.85**	3.97 **7.46**	3.87 **7.19**	3.73 **6.84**	3.57 **6.47**	3.41 **6.07**	3.23 **5.65**
8	5.32 **11.26**	4.46 **8.65**	4.07 **7.59**	3.84 **7.01**	3.69 **6.63**	3.58 **6.37**	3.41 **6.03**	3.28 **5.67**	3.12 **5.28**	2.93 **4.86**
9	5.12 **10.56**	4.26 **8.02**	3.86 **6.99**	3.63 **6.42**	3.48 **6.06**	3.37 **5.80**	3.23 **5.47**	3.07 **5.11**	2.90 **4.73**	2.71 **4.31**
10	4.96 **10.04**	4.10 **7.56**	3.71 **6.55**	3.48 **5.99**	3.33 **5.64**	3.22 **5.39**	3.07 **5.06**	2.91 **4.71**	2.74 **4.33**	2.54 **3.91**
11	4.84 **9.65**	3.98 **7.20**	3.59 **6.22**	3.36 **5.67**	3.20 **5.32**	3.09 **5.07**	2.95 **4.74**	2.79 **4.40**	2.61 **4.02**	2.40 **3.60**

TABLE D—Continued

Degrees of freedom for smaller mean square	Degrees of freedom for greater mean square									
	1	2	3	4	5	6	8	12	24	∞
12	4.75 **9.33**	3.88 **6.93**	3.49 **5.95**	3.26 **5.41**	3.11 **5.06**	3.00 **4.82**	2.85 **4.50**	2.69 **4.16**	2.50 **3.78**	2.30 **3.36**
13	4.67 **9.07**	3.80 **6.70**	3.41 **5.74**	3.18 **5.20**	3.02 **4.86**	2.92 **4.62**	2.77 **4.30**	2.60 **3.96**	2.42 **3.59**	2.21 **3.16**
14	4.60 **8.86**	3.74 **6.51**	3.34 **5.56**	3.11 **5.03**	2.96 **4.69**	2.85 **4.46**	2.70 **4.14**	2.53 **3.80**	2.35 **3.43**	2.13 **3.00**
15	4.54 **8.68**	3.68 **6.36**	3.29 **5.42**	3.06 **4.89**	2.90 **4.56**	2.79 **4.32**	2.64 **4.00**	2.48 **3.67**	2.29 **3.29**	2.07 **2.87**
16	4.49 **8.53**	3.63 **6.23**	3.24 **5.29**	3.01 **4.77**	2.85 **4.44**	2.74 **4.20**	2.59 **3.89**	2.42 **3.55**	2.24 **3.18**	2.01 **2.75**
17	4.45 **8.40**	3.59 **6.11**	3.20 **5.18**	2.96 **4.67**	2.81 **4.34**	2.70 **4.10**	2.55 **3.79**	2.38 **3.45**	2.19 **3.08**	1.96 **2.65**
18	4.41 **8.28**	3.55 **6.01**	3.16 **5.09**	2.93 **4.58**	2.77 **4.25**	2.66 **4.01**	2.51 **3.71**	2.34 **3.37**	2.15 **3.01**	1.92 **2.57**
19	4.38 **8.18**	3.52 **5.93**	3.13 **5.01**	2.90 **4.50**	2.74 **4.17**	2.63 **3.94**	2.48 **3.63**	2.31 **3.30**	2.11 **2.92**	1.88 **2.49**
20	4.35 **8.10**	3.49 **5.85**	3.10 **4.94**	2.87 **4.43**	2.71 **4.10**	2.60 **3.87**	2.45 **3.56**	2.28 **3.23**	2.08 **2.86**	1.84 **2.42**
21	4.32 **8.02**	3.47 **5.78**	3.07 **4.87**	2.84 **4.37**	2.68 **4.04**	2.57 **3.81**	2.42 **3.51**	2.25 **3.17**	2.05 **2.80**	1.81 **2.36**
22	4.30 **7.94**	3.44 **5.72**	3.05 **4.82**	2.82 **4.31**	2.66 **3.99**	2.55 **3.75**	2.40 **3.45**	2.23 **3.12**	2.03 **2.75**	1.78 **2.30**
23	4.28 **7.88**	3.42 **5.66**	3.03 **4.76**	2.80 **4.46**	2.64 **3.94**	2.53 **3.71**	2.38 **3.41**	2.20 **3.07**	2.00 **2.70**	1.76 **2.26**

Degrees of freedom of smaller mean square	24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
		7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.36**	**3.03**	**2.66**	**4.21**
	25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
		7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.32**	**2.99**	**2.62**	**2.17**
	26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
		7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.29**	**2.96**	**2.58**	**2.13**
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
		7.68	**5.49**	**4.60**	**4.11**	**3.78**	**3.56**	**3.26**	**2.93**	**2.55**	**2.10**
	28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
		7.64	**5.45**	**4.57**	**4.07**	**3.75**	**3.53**	**3.23**	**2.90**	**2.52**	**2.06**
	29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
		7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.20**	**2.87**	**2.49**	**2.08**
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
		7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.17**	**2.84**	**2.47**	**2.01**
	35	4.12	3.26	2.87	2.64	2.48	2.37	2.22	2.04	1.83	1.57
		7.42	**5.27**	**4.40**	**3.91**	**3.59**	**3.37**	**3.07**	**2.74**	**2.37**	**1.90**
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.52
		7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**2.99**	**2.66**	**2.29**	**1.82**
	45	4.06	3.21	2.81	2.58	2.42	2.31	2.15	1.97	1.76	1.48
		7.23	**5.11**	**4.25**	**3.77**	**3.45**	**3.23**	**2.94**	**2.61**	**2.23**	**1.75**
	50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	1.95	1.74	1.44
		7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.19**	**2.89**	**2.56**	**2.18**	**1.68**
	60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
		7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.82**	**2.50**	**2.12**	**1.60**
	70	3.98	3.13	2.74	2.50	2.35	2.23	2.07	1.89	1.67	1.35
		7.01	**4.92**	**4.07**	**3.60**	**3.29**	**3.07**	**2.78**	**2.45**	**2.07**	**1.53**

TABLE D—Continued

Degrees of freedom for smaller mean square	Degrees of freedom for greater mean square									
	1	*2*	*3*	*4*	*5*	*6*	*8*	*12*	*24*	∞
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.88	1.65	1.31
	6.96	**4.88**	**4.04**	**3.56**	**3.26**	**3.04**	**2.74**	**2.42**	**2.03**	**1.47**
90	3.95	3.10	2.71	2.47	2.32	2.20	2.04	1.86	1.64	1.28
	6.92	**4.85**	**4.01**	**3.53**	**3.23**	**3.01**	**2.72**	**2.39**	**2.00**	**1.43**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.03	1.85	1.63	1.26
	6.90	**4.82**	**3.98**	**3.51**	**3.21**	**2.99**	**2.69**	**2.37**	**1.98**	**1.39**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.01	1.83	1.60	1.21
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.66**	**2.33**	**1.94**	**1.32**
150	3.90	3.06	2.66	2.43	2.27	2.16	2.00	1.82	1.59	1.18
	6.81	**4.75**	**3.91**	**3.45**	**3.14**	**2.92**	**2.63**	**2.31**	**1.92**	**1.27**
200	3.89	3.04	2.65	2.42	2.26	2.14	1.98	1.80	1.57	1.14
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.89**	**2.60**	**2.28**	**1.88**	**1.21**
300	3.87	3.03	2.64	2.41	2.25	2.13	1.97	1.79	1.55	1.10
	6.72	**4.68**	**3.85**	**3.38**	**3.08**	**2.86**	**2.57**	**2.24**	**1.85**	**1.14**
400	3.86	3.02	2.63	2.40	2.24	2.12	1.96	1.78	1.54	1.07
	6.70	**4.66**	**3.83**	**3.37**	**3.06**	**2.85**	**2.56**	**2.23**	**1.84**	**1.11**
500	3.86	3.01	2.62	2.39	2.23	2.11	1.96	1.77	1.54	1.06
	6.69	**4.65**	**3.82**	**3.36**	**3.05**	**2.84**	**2.55**	**2.22**	**1.83**	**1.08**
1000	3.85	3.00	2.61	2.38	2.22	2.10	1.95	2.76	1.53	1.03
	6.66	**4.63**	**3.80**	**3.34**	**3.04**	**2.82**	**2.53**	**2.20**	**1.81**	**1.04**
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.51**	**2.18**	**1.79**	

TABLE E: χ^2 Table. P gives the probability of exceeding the tabulated value of χ^2 for the specified number of degrees of freedom (df). The values of χ^2 are printed in the body of the table.

df	*0.95*	*0.90*	*0.80*	*0.70*	*0.50*	*0.30*	*0.20*	*0.10*	*0.05*	*0.02*	*0.01*
1	0.00393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	1.145	1.610	2.343	3.000	4.351	0.064	7.289	9.236	11.070	13.388	15.086
6	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.083	16.812
7	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.760	27.587	30.995	33.409
18	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

TABLE F: Conversion of a Pearson *r* into a corresponding Fisher's *z* coefficient.*

r	*z*	*r*	*z*	*r*	*z*	*r*	*z*	*r*	*z*	*r*	*z*
.25	.26	.40	.42	.55	.62	.70	.87	.85	1.26	.950	1.83
.26	.27	.41	.44	.56	.63	.71	.89	.86	1.29	.955	1.89
.27	.28	.42	.45	.57	.65	.72	.91	.87	1.33	.960	1.95
.28	.29	.43	.46	.58	.66	.73	.93	.88	1.38	.965	2.01
.29	.30	.44	.47	.59	.68	.74	.95	.89	1.42	.970	2.09
.30	.31	.45	.48	.60	.69	.75	.97	.90	1.47	.975	2.18
.31	.32	.46	.50	.61	.71	.76	1.00	.905	1.50	.980	2.30
.32	.33	.47	.51	.62	.73	.77	1.02	.910	1.53	.985	2.44
.33	.34	.48	.52	.63	.74	.78	1.05	.915	1.56	.990	2.65
.34	.35	.49	.54	.64	.76	.79	1.07	.920	1.59	.995	2.99
.35	.37	.50	.55	.65	.78	.80	1.10	.925	1.62		
.36	.38	.51	.56	.66	.79	.81	1.13	.930	1.66		
.37	.39	.52	.58	.67	.81	.82	1.16	.935	1.70		
.38	[illegible]	.53	.59	.68	.83	.83	1.19	.940	1.74		
.39	[illegible]	.54	.60	.69	.85	.84	1.22	.945	1.78		

**r's* under 25 may be taken as equivalent to *z*'s.

**TABLE G: Table of Critical Values of K_D in the Kolmogorov-Smirnov Two-Sample Test.
(Small Samples)**

N	*One-tailed test*		*Two-tailed test*	
	α=.05	*α=.01*	*α=.05*	*α=.01*
3	3	—	—	—
4	4	—	4	—
5	4	5	5	5
6	5	6	5	6
7	5	6	6	6
8	5	6	6	7
9	6	7	6	7
10	6	7	7	8
11	6	8	7	8
12	6	8	7	8
13	7	8	7	9
14	7	8	8	9
15	7	9	8	9
16	7	9	8	10
17	8	9	8	10

N	One-tailed test		Two-tailed test	
	$\alpha=.05$	$\alpha=.01$	$\alpha=.05$	$\alpha=.01$
18	8	10	9	10
19	8	10	9	10
20	8	10	9	11
21	8	10	9	11
22	9	11	9	11
23	9	11	10	11
24	9	11	10	12
25	9	11	10	12
26	9	11	10	12
27	9	12	10	12
28	10	12	11	13
29	10	12	11	13
30	10	12	11	13
35	11	13	12	
40	11	14	13	

TABLE H: Table of Critical Values of D in the Kolmogorov-Smirnov Two-sample Test.

(Large samples: two-tailed test)

Level of significance	Value of D so large as to call for rejection of null hypothesis at the indicated level of significance where $D = \text{maximum}\,\lvert S_{n1}(X) - S_{n2}(X)\rvert$
.10	$1.22\sqrt{\frac{n_1+n_1}{n_1n_1}}$
.05	$1.36\sqrt{\frac{n_1+n_2}{n_1n_2}}$
.025	$1.48\sqrt{\frac{n_1+n_2}{n_1n_2}}$
.01	$1.63\sqrt{\frac{n_1+n_2}{n_1n_2}}$
.005	$1.73\sqrt{\frac{n_1+n_2}{n_1n_2}}$
.001	$1.95\sqrt{\frac{n_1+n_2}{n_1n_2}}$

TABLE I: Table of Probabilities Associated with Values as Small as Observed Values of x in the Binomial Test.

Given in the body of this table are one tailed probabilities under null hypothesis for the binomial test when $P=Q=\frac{1}{2}$.

N \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	.031	.188	.500	.812	.969	†										
6	.016	.109	.344	.656	.891	.984	†									
7	.008	.062	.227	.500	.773	.938	.992	†								
8	.004	.035	.145	.363	.637	.855	.965	.996	†							
9	.002	.020	.090	.254	.500	.746	.910	.980	.998	†						
10	.001	.011	.055	.172	.377	.623	.828	.945	.989	.999	†					
11		.006	.033	.113	.274	.500	.726	.887	.967	.994	†	†				
12		.003	.019	.073	.194	.387	.613	.806	.927	.981	.997	†	†			
13		.002	.011	.046	.133	.291	.500	.709	.867	.954	.989	.998	†	†		
14		.001	.006	.029	.090	.212	.395	.605	.788	.910	.971	.994	.999	†	†	
15			.004	.018	.059	.151	.304	.500	.696	.849	.941	.982	.996	†	†	†
16			.002	.011	.038	.105	.227	.402	.598	.773	.895	.962	.989	.998	†	†
17			.001	.006	.025	.072	.166	.315	.500	.685	.834	.928	.975	.994	.999	†
18			.001	.004	.015	.048	.119	.240	.407	.593	.760	.881	.952	.985	.996	.999
19				.002	.010	.032	.084	.180	.324	.500	.676	.820	.916	.968	.990	.998
20				.001	.006	.021	.058	.132	.252	.412	.588	.748	.868	.942	.979	.994
21				.001	.004	.013	.039	.095	.192	.332	.500	.668	.808	.905	.961	.987
22					.002	.008	.025	.067	.143	.262	.416	.584	.738	.857	.933	.974
23					.001	.005	.017	.047	.105	.202	.339	.500	.661	.798	.895	.953
24					.001	.003	.011	.032	.076	.154	.271	.419	.581	.729	.846	.924
25						.002	.007	.022	.054	.115	.212	.345	.500	.655	.788	.885

† 1.0 or approximately 1.0.

TABLE J: Table of Critical Values of T in the Wilcoxon Matched-Pairs Signed-Ranks Test.

N	*Level of significance for one-tailed test*		
	·025	·01	·005
	Level of significance for two-tailed test		
	·05	.02	·01
6	0		—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

TABLE K: Table of Critical Values of Pearson Product Correlation at the .05 and .01 Levels of Significance.

Degrees of freedom (N–2)	*.05*	*.01*	*Degrees of freedom (N—2)*	*.05*	*.01*
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.233	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

TABLE L: Table of Critical Values of Spearman Rank Order Correlation at .05 and .01 Levels of Significance (One-tailed test).

N	*.05*	*.01*	*N*	*.05*	*.01*
5	.900	1.000	16	.425	.601
6	.829	.943	18	.399	.564
7	.714	.893	20	.377	.534
8	.643	.833	22	.359	.508
9	.600	.783	24	.343	.485
10	.564	.746	26	.329	.465
12	.506	.712	28	.317	.448
14	.456	.645	30	.306	.432

TABLE M: Values of r_t taken as the Cosine of an Angle.

Angle	*Cosine*	*Angle*	*Cosine*	*Angle*	*Cosine*
0°	1.000	41°	.755	73°	.292
		42	.743	74	.276
5	.996	43	.731	75	.259
		44	.719	76	.242
10	.985	45	.707	77	.225
		46	.695	78	.208
		47	.682	79	.191
15	.966	48	.669	80	·174
16	.961	49	.656		
17	.956	50	.643	81	.156
18	.951			82	.139
19	.946	51	.629	83	.122
20	.940	52	.616	84	.105
		53	.602	85	.087
21	.934	54	.588		
22	.927	55	.574		
23	.921	56	.559	90	.000
24	.914	57	.545		
56	.906	58	.530		
26	.899	59	.515		
27	.891	60	.500		
28	.883				
29	.875	61	.485		
30	.866	62	.469		
		63	.454		
31	.857	64	.438		
32	.848	65	.423		
33	.839	66	.407		
34	.829	67	.391		
35	.819	68	.375		
36	.809	69	.358		
37	.799	70	.342		
38	.788				
39	.777	71	.326		
40	.766	72	.309		

S.C.E.R.T. West Bengal
Library
Calcutta
25/3, B. C. Road

LOKESH KOUL earned his Ph.D
Education from Kurukshetra Uni
sity after obtaining Post-Gradu
Degrees in Mathematics and E
cation. After serving the Dep
ment of Education, Kuruksh
University, he jointed Departn
of Education, Himachal Prad
University, Shimla as Assoc
Professor. Presently, he is worl
as Professor, and Dean, Faculty
Education, Himachal Pradesh
versity, Shimla.

Dr Koul has contributed a n
ber of research papers to the var
journals of repute, and has gu
a number of Ph.D. and M.
dissertations. He has also dire
NCERT and UGC research Pro
on Tribal and other disadvanta
groups of school going childre
Himachal Pradesh.

Dr Koul has served as U.
Visiting Fellow and delivered a s
of lectures on issues and prob
of educational research at var
Universities in the country. He
also served as a Member of U.
Panel on Education and Membe
the U.G.C. Curriculum Develop
Committee in Education. Prese
he is associated with the E
tional Research and Innova
Committee (ERIC) of NCERT
is the Co-ordinator of Depart
of Special Assistance (DSA)
ramme of the University G
Commission for the Departme
Education, Himachal Pradesh
versity, Shimla.

ISBN 0-7069-4266-3

VIKAS PUBLISHING HOUSE PVT LTD